This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No. 875299

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

D8.1: Report on the link between IDMP and Pharmacotherapeutic Groups and the Need for Medical Data in Pharmacotherapeutic Audit (M22)

Version:	1.0
Status:	Final
Dissemination Level[1]:	PU
Due date of deliverable:	30.09.2021
Actual submission date:	30.09.2021
Work Package:	WP8: **IDMP for Clinical Care, Patients, Pharmacies, Research and Pharmacovigilance**
Lead partner for this deliverable:	I-HD
Partner(s) contributing:	DWIZ, FOUND, SNOMED, CEB/CBG, HAR
Deliverable type[2]:	R

Main author(s):

Vander Stichele Robert I-HD

Other author(s):

Dipak Kalra	I-HD	Anja Van Haren	CEB/CBG
Lucia Comnes	Datawizard	Yuri Quintana	Harvard
Maurizio Taglialatela	Found	Joseph Roumier	I-HD
Jane Millar	Snomed International	Dirk Van Nimwegen	I-HD

[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	05.07.2021	Table of content	Robert Vander Stichele I-HD
0.2	01/O9/21	Revised table of content	Robert Vander Stichele I-HD
0.3	O5/09/21	First 5 chapters	Robert Vander Stichele I-HD
0.4	09/09/21	First inner circle draft	Robert Vander Stichele, I-HD
0.5	14/09/2021	Draft for internal review	Robert Vander Stichele,Dipak Kalra, Josepth Roumier, Dirk Vannimwegen; I-HD
0.6	16/09/21	Internally reviewed version	Luc Nicolas
Final	29/09/2021	Final version	Robert Vander Stichele

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

In this deliverable, crucial steps are reported in linking precise global identification of **individual medicinal products** from different countries (using the IDMP standard) to **international drug classifications**, such as the World Health Organisation Anatomical Chemical Therapeutic Classification (ATC), SNOMED-CT, RxNorm, and other simplified educational classifications. The clinical value of such linking requires to be demonstrated. Ir will be illustrated by showing the usefulness of a link between the global identification number (the Pharmaceutical Product Identifier or PHPID) and the ATC classification. This link will then facilitate the application of internationally validated rules for decision support in pharmacotherapeutic audit.

The report provides a detailed analysis of the coding systems for substances, dose form (EDQM and RxNorm), and the possible business rules for representing strength according to the granularity of substances, and the patterns of dose form. This work was set up in preparation of establishing a procedure for the production of the Pharmaceutical Product Identifier (PHPID), with codification rules for substance, dose form and strength, fed into a Hash function, to produce a global unique identification number for a group of identical medicinal products from different countries, sharing the same 3 items. A public repository of such PhPIDs would be instrumental for applications in clinical care, research, patient information, pharmacovigilance, and precision medicine.

Keywords: Substance, Dose Form, Strength, Pharmaceutical Product Identifier, International Drug Classifications, Pharmacotherapeutic Audit, (In)Appropriate Prescribing

TABLE OF CONTENTS

List of abbreviations

Abbreviation	Complete form
ATC	Anatomical Therapeutic Chemical Classification
CAS	Chemical Abstract Services
CD Precisely	Clinical Drug Precisely
CSD	Clinical Semantic Drug
DDD	Defined Daily Dose
DID	Defined Daily Doses per 1000 inhabitants per day
Dm+d	Dictionary of Medicines and Devices (UK)
EDQM	European Directorate for the Quality of Medicines
EMA	European Medicines Agency
EU	European Union
EU-SRS	European Substance Registration System (EMA)
EUTCT	European Union Telematics Controlled Terms
FDA	Food and Drug Administration
GSRS	Global Substance Registration System (FDA)
ICD	International Classification of Diseases
ICPC	International Classification of Primary Care
IDMP	Identification of Medicinal Products suite of ISO Standards
INN	International Non-Proprietary Name
INNM	International Non-proprietary Name Modified
LOD	Linked Open Data
MeSH	Medical Subject Headings
MOH	Ministry of Health
MP	Medicinal Product
MP only	Medicinal Product Only (SNOMED-CT)
NDA	New Drug Application
NLM	National Library of Medicine
OHDSI	“Observational Health Data Sciences and Informatics” consortium

OMOP-CDM	Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM)
PHPID	Pharmaceutical Product Identifier
PIM-list	Potentially Inappropriate Medication-list
PREMs	Patient Reported Experience Measures
PROMs	Patent Reported Outcome Measures
SAM	Source Authentique de Médicaments (BE)
SDO	Standard Developing Organization
SNOMED-CT	Systematized Nomenclature of Medicine -- Clinical Terms
SPOR	Substance/Product/Organisation/References services
UMLS	Unified Medical Language System
UNII	Unique Ingredient Identifier
VMP	Virtual Medicinal Product
VMPGroup	Virtual Medicinal Product Group
VTM	Virtual Therapeutic Moiety
WHO_UMC	World Health Organisation Uppsala Monitoring Centre

List of figures

List of tables

List of annexes

Annex 1. List of Medicinal products from BE for the 35 substances from the UNICOM Pilot Product List

Annex 2. Result of data cleansing of substances on the Unicom Pilot Product List

Annex 3. Report of theWHO-UMC/FDA Pilot for PhPID production

Annex 4. Analysis of EDQM terminology

Annex 5. Comparison EDQM / RxNorm

Annex 6. Specifications for the Draft Procedure for PhPID production

1 Executive summary

In this deliverable, we report crucial steps in linking precise global identification of individual medicinal products from different countries (using the IDMP standard) to international drug classifications, such as the World Health Organisation Anatomical Chemical Therapeutic Classification (ATC), SNOMED-CT, RxNorm, and simplified educational classifications. The clinical value of such linking is illustrated by applying this link to internationally validated rules for decision support in pharmacotherapeutic audit of the quality of prescribing to older adults in nursing homes.

The kernel concept in IDMP is the Pharmaceutical Product Identifier (PHPID). It is a coding system based on global codification rules for substance, dose form and strength, fed into a Hash function, to produce a global unique identification number for a group of identical medicinal products from different countries, sharing the same 3 items. In this report, we focus on the PHPID_Level IV, which brings together substance, dose form, and strength.

We made an analysis of the coding systems for substance (INN, EUTCT, UNII, CAS, and SNOMED-CT), of the coding system for Dose form (EDQM, SNOMED-CT, RxNORM), and of the business rules to represent the strength of medicinal products, according to the correct granularity of substance and type of dose form. We conducted an analysis was of the EDQM dose form terminology, with several comments and suggestions for improvement. We created a small ontology for substance (moiety, modified substance, International Non Proprietary Name). and for dose form, to create more aggregate groupings of similar PHPIDs, suitable for INN Prescribing and substitution, including a link to the ATC-Level 5 codes, and the other international classifications. We also provide an overview of the results of the WHO_UMC/FDA pilot on the procedure for production of PhPIDs.

We provided a functional analysis of a Linked Open Data Repository of draft PhPIDs to represent all this information, and to allow the recording of correct linking to international drug classifications.

We have developed a short inventory of international classification systems with an approach to link them to national medicinal product dictionaries, based on the use of PhPID, principles of INN Prescribing, and ATC.

In this deliverable, we describe results from the Unicom Pilot Product List. It will be used to populate the repository with IDMP information relevant to PhPID production for the 35 active substances of the UNICOM Pilot Product List, and for the Covid-19 vaccines.

We also describe future work to be performed to connect algorithms in decision rules to the classification systems, to PhPIDs and ultimately to national medicinal product dictionaries.

We finally provide a first approach to possible applications of the PhPID for the control of quality of medical data in the International Patient Summary and Electronic Health Record.

2 Rationale for the use of the ISO/CEN IDMP Standards in clinical care

2.1 Introduction to the use of IDMP in clinical care

In clinical care, physicians, pharmacists and nurses deal with medicinal products all the time. Pharmacotherapy is a frequent and important intervention at the disposal of these healthcare workers.

Physicians make diagnoses or develop hypotheses about the ailment of their patients. Then they decide whether or not the ailment can be addressed with pharmacotherapy, and if so, they must scan the pharmacotherapeutic arsenal to choose the right medication.(Denig & Haaijer-Ruskamp,1992); (Denig, 1994) According to the WHO, rational use of medicines requires that patients receive medications appropriate to their clinical needs, ***in doses that meet their own individual requirements, for an adequate period of time, and at the lowest cost to them and their community***. (de Vries et al., 1995; Henning et al., 2001)

To be able to make the right choice, according to that definition, physicians and pharmacists must be well aware of the medicinal products, authorized in their jurisdiction. They must know the branded originator products and the (branded) generic medications, with their substances, dose forms, strengths, prices, pack sizes, and reimbursement status.

The intricate integration of knowledge of pharmacotherapeutic classes and the details of medicinal products is a delicate balance.(Aronson & Aronson, 2012; Maxwell A, 2016) It requires support from medicinal product dictionaries(MPD) that provide physician desk reference services in printed and on line versions. Computerised decision support systems (CDSS) can provide instant alerts during each act of prescribing or support periodic reflexion in the process of medication review of poly-medicated patients.

Building adequate medicinal product dictionaries and decision support systems requires a deep ***understanding of the drug choice process*** of the health providers.

The World Health Organisation has provided an educational philosophy and materials to teach the art and science of prescribing and to promote good prescribing, clearly embedded in an approach of evidence-based medicine (Henning et al., 2001); (Van Doorn et al., 2009); (Tichelaar et al., 2020).

All this requires an accurate method of identification and description of the medicinal products, preferably in a global approach.

The ISO/CEN standards for the identification of medicinal products (IDMP) provide a basis to reorganise regulatory information on medicinal products, as originated in pharmaceutical companies, and validated by regulatory authorities. The standards assure the interoperability of information on medical products in information systems, from the industry, the regulator, the medicinal product dictionary providers, vendors of electronic health care records, developers of apps for patients.

The UNICOM Action Programme was designed to support the implementation of these standards in this chain of information providers, starting with the industry and the regulators. However, the intent was clearly present to make sure that this renewed and standardised way of presenting drug information would also benefit clinical care and research in pharmaco-epidemiology.

It is hoped that the standardised way to identify and describe medicinal products will find its way into the medicinal product databases of agencies, drug information centres, and of international and national publishers. From there the information can trickle down into the electronic prescribing systems, the electronic health records, and the databases recording drug use and reimbursement.

IDMP can so become an essential part in the machinery of drug information, by providing standardised identification and description of medicinal products. It can assure interoperability between the systems of research and development departments of pharmaceutical companies, marketing authorisation agencies, pharmacovigilance centres, drug information centres, medicinal product dictionaries, decision support publishers, vendors of electronic health care systems, and real-world data managers.

2.2 The need to precisely identify medicinal products within one country

Between medicinal products with the same granular substance, dose form, and strength, slight variations might still exist. There are brand originators, co-marketing drugs and branded generics, marketed by different marketing authorisation holders. There might be subtle differences in ingredients (with or without clinical interest). Details of dose form may vary (e.g. shape, colour, taste, ease of intake, imprints of tablets, division marks (with or without guarantee that the content in active substance is divided equally). Whether dose forms might be split, crushed or mixed may vary. Prices or reimbursement rules can vary substantially. The list of authorized indications can be different between brands and generics. Pack sizes, secondary and immediate containers may vary (e.g. Tablets in a blister pack in folding carton versus tablets in a glass container).

The aim of the IDMP ISO-standards is to standardize the detailed description of a medicinal product, not only its active substance(s), dose form and strength, but ***also the other details***, which may or may not be important for clinical decision support and drug choice. Beyond IDMP it is important that the clinical particulars (indications, dosing information, contra-indications, unwanted effects, etc..) in the textual information of the drug labelling becomes structured and standardized for each authorized medicinal product in the jurisdiction. This standardisation of the labelling should trigger interoperability between the information systems of the marketing authorisation agency, the Ministries of Health (MOH) , health insurance institutes, eHealth, and EHR-vendor systems in hospital and primary care. Only then can drug information seamlessly flow in the health eco-systems, and correctly and practically reach the healthcare providers and patients.

2.3 The need to precisely identify similar medicinal products from different jurisdictions

2.3.1 Cross-border prescribing and dispensing

In the past decade, several European initiatives have contributed to create the preconditions for cross-border electronic prescribing and dispensing, within Europe, and between Europe and the US (EpSOS, OpenMedicine, Trillium Bridge I and II). Making this happen has become a priority for the eHealth systems of the member states, and has been a priority for the European eHealth Network. It is the main objective of Work Package 5 to 7 in the Unicom Action Programme.

The implementation of IDMP in the different national Medicinal Product Dictionaries is a cornerstone of this endeavour, as it will permit (in most cases) that a prescription for a medicinal product in one country can be received and interpreted, and finally lead to the dispensing of an (almost) identical medicinal product in another country. For a correct dispensing process, it is important that also an interpretable version of the full medication list is provided as context for the prescription, as well as vital clinical information, as contained in the International Patient Summary. All this requires complicated electronic infrastructure, accompanying regulation, interoperable systems, global identification of medicinal products with PHPIDs, and multilingual descriptions of the essential characteristics of pharmaceutical products (substance, dose form, and strength).

Cross-border ePrescription may currently be an exceptional scenario, but may become much more frequent, in the light of intensified travel within Europa, and between Europe and other continents. It is caused by increasing international business, tourism, work, intensified integration of regions from neighbouring member states, and medical cooperation between hospitals from different countries. This use case will be instrumental to illustrate to Member States the importance of interoperability issues.

2.3.2 Evidence-based Medicine crossing the borders

However, it is not only the prescription that needs to cross borders. Also, the information about the medication self must be able to pass smoothly between member states and countries. In the age of Evidence-Based Medicine, scientific knowledge is translated into guidelines and into decision support systems, which contain specific alerts and specific practice recommendations, triggered by specific clinical situations as recorded in the Electronic Healthcare Record (EHR) systems. Recently, a multitude of mobile applications for healthcare providers and for patients have been createdThis illustrates the

growing importance of the citizen as actor in the health care process, providing data on preferences and outcomes (PREMS and PROMS).

The European Common Market counts 27 member states and 24 official languages. Each country has one to several medicinal product dictionaries. The developers of decision support systems and mobile apps face a fragmented market, must translate their content, and adapt to local drug dictionaries, which are all to updated at a frantic rhythm. It makes it hard and expensive for publishers to build competitive drug information systems, in comparison to a vast single market in the USA, with English as the dominant language.

Europe harbours a lot of high-level medical university centres, developing word class applications in very specific areas (drug-drug interaction, pregnancy and lactation information, indication and dosing support, toxicological and pharmaco-genetic information, etc...) These specialised centres can become part of a larger information ecosystem, that will make these centres of expertise sustainable. However, their economical business model depends on ***smooth access to the markets of the member states, translation support, and smooth access to interoperable national medicinal product dictionaries***. Then, and only then, evidence will know no borders.

3 Building a solid foundation for international grouping: The Pharmaceutical Product Identifier (PHPID)

3.1 Rationale

IDMP provides the means to accurately identify and describe each individual medicinal product in a given jurisdiction (e.g. a member state of the European Union, the USA, Japan).

It is preferable that the elements to identify and describe medicinal products are standardized, which means that the list of elements used is the same in all jurisdictions, and that these elements (and their value sets) are expressed by terms from controlled multilingual vocabularies and reliable coding systems.

The more accurate the description of individual medicinal products is made, the more reliable the international grouping of (almost (identical) medicinal products on the basis of one or more characteristics can be performed. ***The possibilities of groupings are endless, but some are more relevant than others.***

3.2 The 3 crucial elements of the Pharmaceutical Product Identifier

Within IDMP, a grouping on the basis of 3 crucial variables is established:

1. The Substance (the modified substance in case of chemicals)
2. The Dose Form (more specifically the administrable dose form)
3. The Strength (of the administrable dose form)

With global, universally accepted codes for each of these 3 elements, a unique identifier can be created which combines these 3 elements, using a HASH function.

In IDMP this unique identifier of a group of medicinal products, containing the same 3 elements is called the PhPID (Pharmaceutical Product identifier).

There are 4 levels in this identifier:

1. PhPID_L1: Only substance
2. PhPID_L2: Substance + Strength (little practical value)
3. PhPID-L3: Substance + Dose Form
4. PhPID-L4: Substance + Dose Form + Strength

Concepts similar to PHPID-L4 already exist in **RxNorm** (Semantic Clinical Drug), in **Snomed-CT** (Clinical Drug – CD Precisely), in the UK Dictionary of Medicines and Devices **-Dm+d** - (Virtual Medicinal Product -VMP), albeit the granularity of the value set of substance and of dose form, and the way of expressing strength might slightly differ between these information systems.

3.3 Two major applications relevant for clinical care

With these 3 basic elements (substance, granular dose form, and referenced strength, two major applications can be realised:

1. a solid operationalisation of **INN prescribing** (prescribing by International Non-Proprietary Name) can be facilitated,
2. links can be established with
    - Classes for drug utilisation research in the taxonomy of the WHO Anatomical Chemical Therapeutic Classification,

- SNOMED-CT drug Classes,
- WHODRUG Standardized Drug Classes,
- RxNorm (and the common data model OMOP from OHDSI),
- the table of content of pharmacological handbooks and pharmacopoeias
- the drug classes defined in decision rules of decision support systems.

3.4 Focus on PHPID in Work Package 8 of the UNICOM project

In this deliverable, the focus is on the grouper PhPID (the Pharmaceutical Product Identifier).

This will provide the solid foundation for sensible, robust, and precise grouping of medicinal products across jurisdictions. It is based on uniform description of substance, dose form and strength in an abstraction of (almost) identical medicinal products from different jurisdictions, independent of the country and the marketing authorisation holder.

This basic grouping can be the starting point for further clinical grouping of these atomic PhPID groups, guaranteeing correct information of the 3 basic constituents (substance, dose form, strength), needed for a correct connection to higher pharmaco-therapeutic classes.

Because precision is needed for this basic grouping, the three constituent elements must be well defined:

- the substance needs to be described at its most relevant granularity level (meaning the modified active substance for ionized chemicals)
- the dose form should be at the most relevant granular level for administrable dose forms, and supported by solid characteristics and definitions (as in EDQM and in SNOMED-CT).
- The strength needs to be a standardized strength, with a Basis of Substance Strength (BoSS) that is the same for all members of the group.

This precision may sometimes be too far reached for clinical purposes, but is needed for regulatory purposes and pharmacovigilance use cases. It provides the basis for solid processes of aggregation to higher levels of abstraction, which might be more relevant for clinical care use cases.

- *Note : For more specific groupings, based on more intricate details of the medicinal products, open access to the standardized drug databases of the jurisdiction will be needed, to provide details, such as presence of inactive ingredients of clinical interest, other inactive ingredients, shape, colour, taste, range of permitted routes of administration, pack size, as well as regulatory information such as belonging to controlled substances, Over the Counter (OTC) or on prescription (Rx), subject to reimbursement rules, prices, etc. For a description of what is needed beyond this basic identification to run adequate decision support systems, we recommend a publication of the pharmacological department of Heidelberg University, Germany.* (Senger et al., 2011)

3.5 Responsibility for PhPID production

The exact and standardized determination of the 3 main elements in the identification of medicinal products is of course the responsibility of each local jurisdiction. However***, super-national governance and validation process of these crucial attributions will be needed***, if interoperability for this basic and pivotal concept is to be achieved.

Increased cooperation between industry, agencies, and institutions such as the WHO Uppsala Monitoring Centre for Pharmacovigilance can ensure a smooth validation of this basic identification of any (new and old) medicinal product (see further). Once a limited number of countries have identified and validated the PhPID of all their medicinal products, and provided that this information becomes publicly available, it will be much easier for other agencies to follow the same pathway.

4 Moving to pharmacotherapeutic groups, by gradually building groups of higher level of abstraction.

4.1 Starting with the level of PhPID

The Pharmaceutical Product Identifier (PhPID) groups medicinal products from different countries and different companies that share the same substance(s), dose form and strength. These 3 elements are expressed in standardized terms from controlled terminologies, at a high level of granularity (modified substance, when available; granular dose form (EDQM or SNOMED-CT); and strength.

This triplet of elements pulls together the (almost) identical medicinal products of different countries and companies, and provides a solid foundation for further aggregation. In this deliverable, we will use the example of the medicine amlodipine (for the treatment of hypertension and angina pectoris), and for illustrations with authorized medicinal products, we will use products from the Belgian Market, as these data are readily and publicly available in the SAM (Source Authentique de Médicaments) database of the eHEALTH system.

4.1.1 Identifying the triplets for amlodipine as an example

As an example, Table 1 lists the available combinations of substance / dose form/ strength for medicinal products containing only the calcium antagonist amlodipine (so not combination products). This medication is marketed in 3 modified substances (amlodipine besylate, amlodipine mesylate, amlodipine maleate). Hence all these products contain the single moiety “amlodipine”, but with different modifiers. There are 2 strengths (5 mg and 10 mg) and three different dose forms available (capsule,hard; tablet; coated tablet). That leads to 9 different triplets.

Looking at the data for Belgium, we can see that there is amlodipine besylate / capsule, hard / 5 mg and amlodipine besylate / capsule hard /10 mg (both represented with 1 brand, marketed by the originator). No other company markets amlodipine in this dose form. There are two other triplets with amlodipine besylate / tablet, one with 5 mg (marketed by 6 generic companies) and one with 10 mg (marketed by 7 generic companies). Triplets with amlodipine mesylate are not (anymore) available in Belgium. Triplets with amlodipine maleate come in tablets for 5 mg, and in tablets and coated tablets for 10 mg;one company has a coated table that can be split (a divule) of 10 mg, which can easily be broken in two halves. Most companies bring their medicinal products in a small (28 to 30 pack units) and a large pack (98 to 100 pack units). (See Table 1 for an overview). In **Annex 1**, the complete list of medicinal product packs in Belgium is given for the 35 substances of the UNICOM Pilot Product List).

Table 1. Overview of Pharmaceutical Product Groups in the Belgian national market), containing the single moiety amlodipine.

Pharmaceutical Product Group (virtual medicinal product)				
			Data from Belgium	
Modified substance	Granular dose form	Strength	Orginator Company	Generic Companies
amlodipine besilate	capsule, hard	5 mg	1	0
amlodipine besilate	tablet	5 mg	0	6
amlodipine mesilate	tablet	5 mg	0	0
amlodpine maleaat	tablet	5 mg	0	1
amlodipine besilate	capsule, hard	10 mg	1	0
amlodipine besilate	tablet	10 mg	0	5
amlodipine mesilate	tablet	10 mg	0	0
amlodipine maleate	tablet	10 mg	0	1
amlodipine maleate	coated tablet	10 mg	0	1
Note: There is 8 active companies in Belgium (1 orginator, 7 generic companies)				
There are 16 medicinal Products and 28 Medicinal Product Packages available for amlopdine				
These belong to 7 different Pharmaceutical Product Groups, each to be defined by a PhPID				
For the prescriber there are 2 basic options: amlopdipine oral 5mg and amlopine oral 10 mg				

In the UNICOM Pilot Product List small team, a comparison was undertaken with the UK, the Netherlands, and Germany, with respect for the non-public character of the drug data in the Netherlands and Germany.

In the UK there are about 40 medicinal product packs (all oral forms and 5 or 10 mg) with 1 originator and 19 generic companies (data with courtesy from Julie James). In the Netherlands, there are 30 medicinal products (number of packs not clear) from 1 originator and 9 generic companies, including unlicensed paediatric doses (1 mg) and liquid dose forms (data with courtesy from Leonora Grandia). In Germany, there are 55 medicinal product packs, from 1 originator and 29 generic companies; one company licenses a 7.5 mg strength (data with courtesy from Ursula Tschorn).

Extending this analysis to these and other countries will lead to a further (but moderate) increase of number of triplets needed to represent amlodipine.

4.1.2 Looking up the available coding systems

Once the available triplets for a given substance are determined, it is possible to look for the exact codes, that represent each of the three basic elements. Coding systems are needed for the modified substances or inert moieties, for the dose forms, and for the units of measurement of strength.

For “modified substance” at least 5 coding systems are available WHODRUG, EUTCT, UNII, CAS, and SNOMED-CT.

For dose form, 3 coding systems are available:EDQM, SNOMED-CT, and RxNorm

For strength only one, widely recognised coding system is available, namely UCUM (Unified Code for Units of Measure, from the Regenstrief Institute.

4.1.3 Looking up available codes for modified substances

In Table 2, we provide an overview of the coding numbers for the 3 modified substances of amlodipine.

It remains to be seen if one of these codes can be the best candidate for a global solution or whether a new global code needs to be defined. It is not clear yet, whether through the SPOR system a European code will be imposed in Europe (either the EUTCT or a new SPOR-code for substance***). In this stage of the development of IDMP implementation this might be problematic***.

Table 2. Available codes for the 3 modified substances of amlodipine.

Coding possibilities for amlodipine					
Modified substances					
	WHODRUG INNM	**EUTCT**	**UNII**	**CAS**	**SNOMED-CT**
amlodipine besylate	00972401001	100000090079	864V2Q084H	111470-99-6	84976003
amlodipine mesylate	00972404001	100000089571	291Y33EZHA	246852-12-0	not present
amlodipine maleate	00972403001	100000089370	CQ27G2BZJM	88150-47-4	421048000

Note:

- *In EUTCT the label is amlodipine besilate and amlodipine mesilate (with besylate and mesylate as synonyms).*
- *For Amlodipine mesylate, very few pharmacovigilance reports, indicating very limited commercial activity.*
- *Theoretically, there is also amlodipine benzoate, but not marketed in the EU (and hence, no pharmacovigilance reports)*
- *It must be noted that the codes in the WHODRUG Dictionary and the terms in the INN vocabulary are currently not connected.*

4.1.4 Looking up avaible codes for dose forms

It is clear that the IDMP standard promotes the use of EDQM terminology for dose forms. Table 3 lists the available codes for 3 dose forms of amlodipine. Other coding systems, such as SNOMED-CT and RxNorm also provide coding and terminology, but both systems would have only one entry to describe the 3 EDQM dose forms listed here, namely: Oral Tablet. There has been ***a proposal from the FDA to use the characteristics of the EDQM terminology as the basis for a new, global coding system***.

Table 3. Available codes for 3 dose forms of amlodipine.

Granular Dose form	
	EDQM
capsule, hard	10210000
tablet	10219000
coated tablet	10220000

4.1.5 Looking up available codes for strength Units

In the UCUM system, the code for mg is “408”. (table 4)

Table 4. Ucum code for milligram

Unit of strength	
	UCUM
mg	408

4.1.6 Completion of collection of codes in preparation of PHPID production

With this approach all elements for starting a HASH function that will result in an PhPID identifier are gathered. This work will have to be completed by looking at the medicinal products that contain amlodipine in combination with other substances (combination preparations). Similar work must be performed for each of the 35 selected substances of the UNICOM Pilot Product List, and ultimately for all medicinal products from the national drug dictionary, country by country.

It may come as a surprise that at this stage of the development of the implementation of IDMP on a global scale, ***crucial issues of uncertainty about the procedure to follow in the production of the Pharmaceutical Product Identifiers are still not settled, neither in the EMA IDMP Implementation Guide, neither within UNICOM, neither at the global level.*** **Intense concertation on the global level, but also experience in pragmatic pilots, facilitated by UNICOM, and maybe coordinated by WHO-UMC, may be useful to achieve consensus on this matter.**

4.2 Building up to the level of Virtual Medicinal Product Group (VMPGroup)

The group of medicinal products with the same PhPID is defined by the modified substance (INNModified) for most of the chemical substances, the most important type of substances in pharmacology.[3] For clinical use cases, this may not always be relevant. Some may argue that there are clinical differences between the different salts of amlodipine, while others may consider their therapeutic activity equivalent, making it irrelevant to know which of the salts is used.

The prescriber may accurately express his/her therapeutic intent by specifying that the pharmacist must dispense

- a medicinal product containing amlodipine (without specifying the salt),
- an oral dose form
- choosing the 5 mg strength, rather than the 10 mg

[3] For a full discussion of the characterisation of different types of substances, see the documents prepared for WG6 of ISO, by Jean-Jacques Gonzages and Herman Diederik.

In Table 5, the possible expressions for these choices are given.

Table 5 . Triplets of higher aggreation for amlodipine

Virtual Medicinal Product Group		
INNsubstance	**Intended Site**	**Strength**
amlopidine	oral	5 mg
amlodipine	oral	10 mg

Further details (e.g. tablet that can be split or not, coated tablet or not) can be specified or left to the discretion of the pharmacist, in dialogue with the patients.

This grouping might also provide the basis for **t*ransparent and fair business rules for substitution regulation***.

In Belgium the principles for establishing these VMPGroups for INN Prescribing have been operationalised in regulation from the agency. (Van Bever et al., 2014)

Again, it is quite possible to gather the codes from official terminologies (in casu WHO for the label of the INN, EDQM for Intended Site, and UCUM for strength, in preparation for a hash function to produce the identifiers for these concepts.

For dose form, it is proposed to use the high-level categorisation of the intended site characteristic of EDQM, rather than the very granular EDQMP dose form. It is possible to create intermediate classes (see further) to assist the prescriber in adequate choices with regard to dose form. Snomed-CT has a similar concept as attribute to the dose form.

4.3 Building up to the level of Virtual Therapeutic Moiety Level (VTM)

Finally, it is possible to make abstraction of the dose form and strength, and only look at the INNsubstance as a criterion to group substances and medicinal products. This grouper concept collects all the available modified substances.

Table 6 provides the possible coding for this concept, in the WHODRUG-INN coding of the INN nomenclature.

Table 6. Potential code for the grouper of modified substances of amlodipine.

INNSubstance	
	WHODRUG-INN
amlodipine	00972401001

Using this concept, we can now construct a collection of medicinal products that all contain this INNsubstance (hence any form of modified amlodipine) in medicinal products with amlodipine as a single substance or in combination products, where amlodipine is one of the constituents.

In Snomed-CT, the most related concept would be "***Medicinal Product ONLY or MP ONLY***". According to the Snomed-CT definition: "MP ONLY is: an abstract representation of a medicinal product based on description of only and exclusively the active ingredient substance(s) that it contains but regardless of any modification of those active ingredient substance(s).

The definition indicates that this concept and other related concepts are meant to make an abstract description of individual medicinal products and is part of the international release (not the national extension). (SNOMED DRUG MODEL description).

> *Note: For the purpose of naming groups of medicinal products by their substances, SNOMED-CT also uses a general concept, called "structural grouper", defined as: A concept grouping together medicinal products based on the chemical structure of their active ingredient nnsubstance(s).*

This level is very close to ***the fifth level of the ATC classification***, which is one way to standardize the concept, serving as an entry to the 5-level taxonomy for Drug Utilisation Research of the World Health Organisation (see further).

Situated at the same level is also the ***European Union Reference Dates (EURD) list***, that governs the coordination of Periodic Safety Update Reports (PSURs), by assigning to each active substance or combination of active substances a responsible agency and a responsible company, at the European level, with the task to periodically review on fixed dates the safety reporting of the therapeutic arsenal.

Note:

- *PhPID_L1 is also an aggregation of medicinal products but characterised by the modified substance (if pertinent). Its equivalent in Snomed-CT is Medicinal Product Precisely, defined as: an abstract representation of a medicinal product based on description of only and exclusively the precise active ingredients it contains*
- *Snomed-CT also has the concept Medicinal Product Form (as ONLY and as CONTAINING), which is intended as a combination of substance and form.*
- *Another concept is the "therapeutic role grouper", defined as: a concept grouping together medicinal products based on a broad description of their use in treatment of disease. That is a very general concept, but laying the bridge to pharmacotherapeutic classification. In addition, the rigorous ontological structure of Snomed-CT allows complex multi-axial classification. For a full discussion of the grouper concept .*
- *The IDMP standard proposes the concept "cluster", a vague and general concept, that is not further elaborated.*

4.4 Dealing with combinations

At this level it becomes important to be able to handle single active substances and combination of active substances, and more importantly the relation between single substances and combinations.

In the examples described up to now, only medicinal products were used, containing a single substance. Yet, in pharmacology, a number of combination products exist. While clinical pharmacologists in general are not in favour of combining more active substances in one medicinal product, the number of combinations is high in ill-regulated markets, with many questionable combinations.(Wirtz et al., 2013) In the past decade, mainly in the field of cardio-vascular primary and secondary prevention, the number of combinations of cardiac medicines has substantially increased, mainly because the scientific associations of cardiologists advocated them.

The case of amlodipine is a good example, as it is present in many combination products (see Table7).

Table 7. Amlodipine single and combinations

Medicinal products with only amlodipine

Amlodipine (single)

amlodipine oral 5 mg
amlodipine oral 10 mg

Medicinal products with amlodipine in combination

ACE inhibitor + Calciumantagonists

Perindopril + amlodipine

perindopril + amlodipine	oral	eq. 4 mg + 5 mg
perindopril + amlodipine	oral	eq. 4 mg + 10 mg.
perindopril + amlodipine	oral	eq. 8 mg + 5 mg
perindopril + amlodipine	oral	eq. 8 mg + 10 mg

Ramipril + amlodipine

ramipril + amlodipine	oral	5 mg + 5 mg
ramipril + amlodipine	oral	5 mg + 10 mg
ramipril + amlodipine	oral	10 mg + 5 mg
ramipril + amlodipine	oral	10 mg + 10 mg

Sartans + Calciumantagonists

Olmesartan + amlodipine

olmesartan + amlodipine	oral	20 mg + 5 mg
olmesartan + amlodipine	oral	40 mg + 5 mg
olmesartan + amlodipine	oral	40 mg + 10 mg

Telmisartan + amlodipine

telmisartan + amlodipine	oral	80 mg + 5 mg
telmisartan + amlodipine	oral	80 mg + 10 mg

Valsartan + amlodipine

valsartan + amlodipine	oral	80 mg + 5 mg
valsartan + amlodipine	oral	160 mg + 5 mg
valsartan + amlodipine	oral	160 m + 10 mg

Ace-inhibitors + Calciumantagonists + Diuretics

Perindopril + amlodipine + indapamine

perindopril + amlodipine + indapamide oral eq.4 mg+5 mg+1,25mg
perindopril + amlodipine + indapamide oral eq. 4 mg + 10 mg+1,25 mg
perindopril + amlodipine + indapamide oral eq. 8 mg + 5 mg + 2,5 mg
perindopril + amlodipine + indapamide oral eq. 8 mg + 10 mg + 2,5 mg

Sartans + Calciumantagonists + Diuretics

Olmesartan + amlodipine +HCT‡

olmesartan + amlodipine + HCT	oral	20 mg + 5 mg + 12,5 mg
olmesartan + amlodipine + HCT	oral	40 mg + 5 mg + 12,5 mg
olmesartan + amlodipine + HCT	oral	40 mg + 5 mg + 25 mg
olmesartan + amlodipine + HCT	oral	40 mg + 10 mg + 12,5 mg
olmesartan + amlodipine + HCT	oral	40 mg + 10 mg + 25 mg.

Valsartan + amlodipine +HCT‡

valsartan + amlodipine + HCT	oral	160 mg + 5 mg + 12,5 mg
valsartan + amlodipine + HCT	oral	160 mg + 5 mg + 25 mg
valsartan + amlodipine + HCT	oral	160 mg + 10 mg + 12,5 mg
valsartan + amlodipine + HCT	oral	160 mg + 10 mg + 25 mg
valsartan + amlodipine + HCT	oral	320 mg + 10 mg + 25 mg

Statins + ACE-Inhibitors + Amlodipine

Atorvastatatine + Perindopril + Amlodipine

atorvastatine + perindopril + amlodipine oral 0 mg + eq. 4 mg + 5 mg
atorvastatine + perindopril + amlodipine oral 20 mg + eq. 4 mg + 5 mg
atorvastatine + perindopril + amlodipine oral 20 mg + eq. 8 mg + 5 mg
atorvastatine + perindopril + amlodipine oral 20 mg + eq. 8 mg +10mg
atorvastatine + perindopril + amlodipine oral 40 mg + eq. 8 mg+10mg

There are several ways to deal with the complexity of combinations. (see figure 1).

One can split a group of medicinal products only containing amlodipine, and another group with combinations all containing amlodipine but specifying the other constituent substances.

One can make a collection of all products containing amlodipine (without specifying the other constituent substances) and with an included subset of medicinal products that only contain amlodipine.

Finally, it is possible to create a database with medicinal products, either single or combinations, but all containing amlodipine, with another database where all single components are listed, with pointers from each component to the relevant medicinal products in the main database).

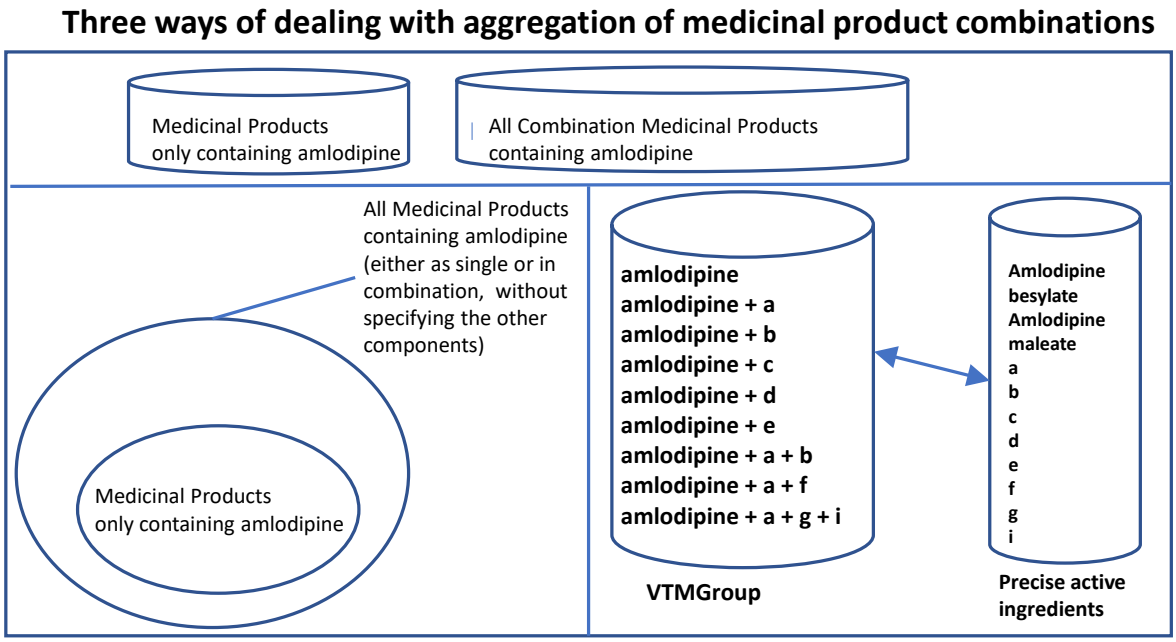


Figure 1. Three ways of dealing with aggregation of combination medicinal products.

4.5 Bringing it all together

Representing medicinal products is the subject of the IDMP Suite of Standards to be implemented, and has been the objective of several operational information systems in the world.

One may cite the pioneering drug model of the UK Dictionary of Medicines and Devices (Dm+d). There is RxNorm in the USA, set up by the National Library of Medicine and used by the FDA and the OSHDI consortium, governing the OMOP Common Data Model.

Snomed-CT has refined its drug data model, with concepts which are part of the international release and hence have a global function, and contents, part of the national extension, for a growing group of countries who want to describe their therapeutic arsenal with Snomed-CT.

In Figure 6, we have tried to align the concepts from these systems, in an overview, which addresses concepts at 3 levels:

1. the national level of concrete medicinal products
2. the first level of abstraction with granular representation of modified substances and detailed dose form
3. the higher level of abstraction with INN-substance and Intended Site characteristic as the basis of aggregation for dose forms.

Note:

- *In RxNorm the value set of dose forms is small and not very detailed. Hence, abstraction is not used in any level of the concepts. In addition, the precise ingredient is immediately abstracted, already at the first level of abstraction.*
- *To find the equivalent of the virtual medicinal product, in Snomed-CT the concept Medicinal Product Form Only has to be combined with the strength.*
- *Both RxNorm and Snomed-CT have a variety of intermediate concepts, involving only substance and dose form, left out in the diagrams, for reasons of clarity.*
- *Both Snomed-CT and RxNorm have concepts to deal with combinations (resp. "Medicinal Product MP containing" and "multiple ingredients".*

In figure 2, an attempt is made to align the concepts for medicinal product pack, medicinal product, and brand name, for the different information systems, and at three levels of abstraction.

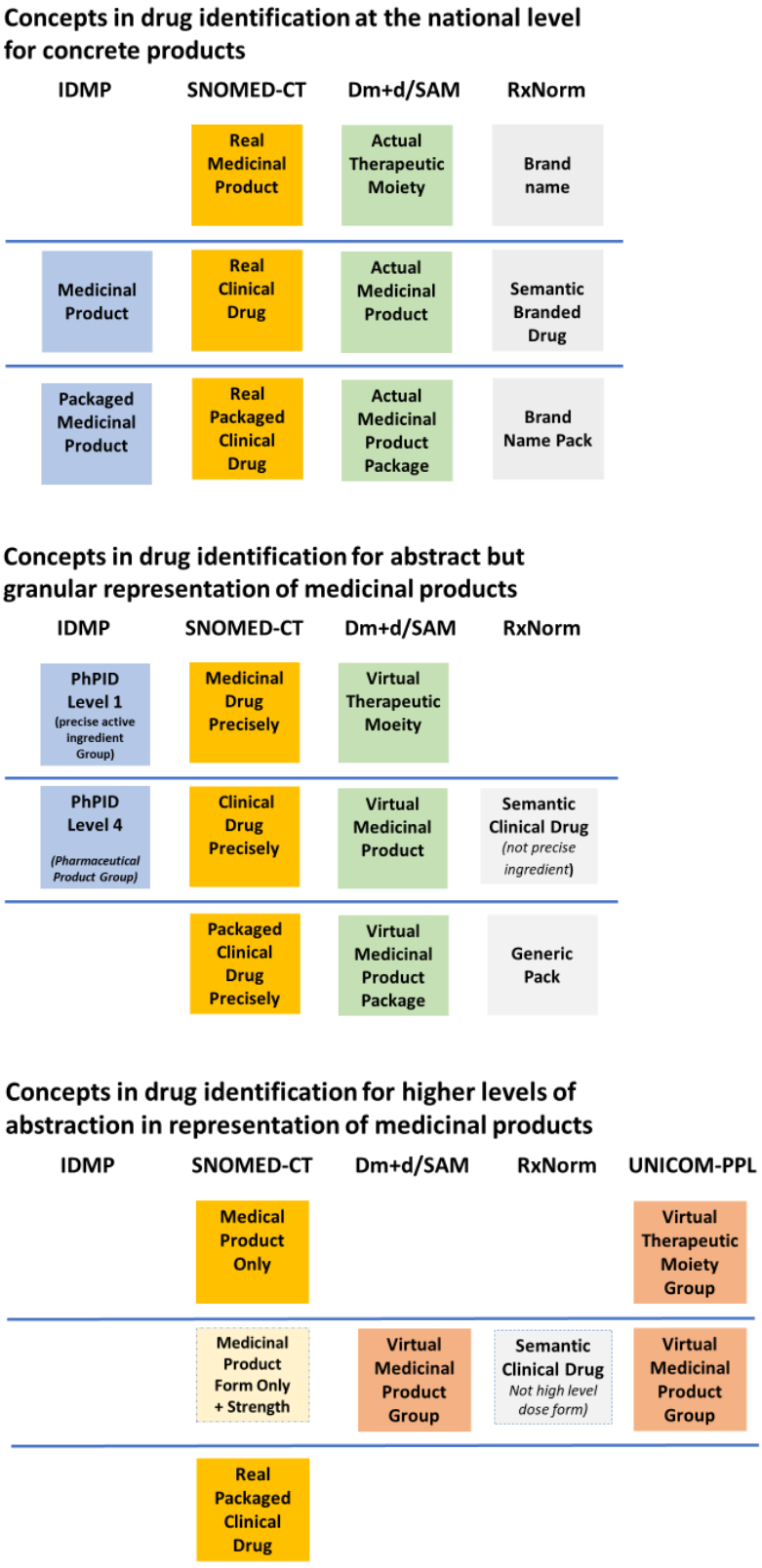


Figure 2. Collation of concepts in representation of medicinal products in different systems

4.6 Moving to pharmacotherapeutic groups

The gradual upgrade in the abstraction of the representation of medicinal products paves the way to the connection with international drug classification.

It then becomes easy to establish a link to the ATC system 5th level (e.g. C08CA01 amlodipine and several other codes for the combination products). For some substances, the strength and the dose form need to be taken into consideration for connecting correctly to the ATC.

The ATC code can be checked against the official code in the labelling. Many countries have linked their entire pharmaceutical arsenal (or only the antibiotics) to the ATC. This made intense drug utilisation research possible and also the mapping of antibiotic consumption in Europe. (Vander Stichele et al., 2004).

From there, the link to the 5-level taxonomy of the ATC classification is self-evident. There are many mappings between the systems (Snomed-CT, RxNorm). (Dhombres & Bodenreider, 2016), (Winnenburg & Bodenreider, 2014).

WHO-UMC, the Uppsala Centre for Pharmacovigilance, has linked the ATC to its system of Standardized Drug Groups[4].

Snomed-CT itself has a rigorous internal ontological approach to allow for multi-axial searching for pharmaceutical classes.

[4] https://www.who-umc.org/whodrug/whodrug-portfolio/whodrug-standardised-drug-groupings-sdgs/

5 Starting to get real. The Unicom Pilot Product List

5.1 Rationale

At the start of the Unicom Project, it became clear that is the implementation of IDMP in national medicinal product dictionaries and in national drug databases, run by the agencies, is in its infancy.

In many Work Packages of UNICOM, there is and will be a need for concrete datasets of medicinal products from multiple countries to be used in deliverables, use cases, and tools, promised in the Description of Action.

Examples are needed:

- To discuss the application of the IDMP standards and implementation priorities with the help of concrete examples
- To illustrate the implementation process of IDMP in National Agencies
- To be used in the cross-border ePrescription and eDispensation pilots
- To illustrate the possibilities of clinical and research applications of IDMP
- To illustrate the implementation process in National Medicinal Product Dictionaries

It was therefore decided early in the upstart of the UNICOM Programme of Action, to build a Unicom Product Pilot List.

5.2 Upstart of the PPL project

Under the initiative of Empirica and Work Package 1, an ad-hoc working group named “Unicom Pilot Product List” was started, as people active in the different work packages of UNICOM perceived an urgent need to proceed at an earlier stage then originally foreseen, with greater speed, and more intensity on this matter.

A small team of experts was created (Julie James, Ursula Tschorn, Leonora Grandia, Robert Vander Stichele) and they together established the principles of constructing a list of 35 substances (and their modifiers), estimated to represent approximate 250 pharmaceutical products (and many more medicinal products and medicinal product packages) from 4 countries (UK, BE, NL, GE).

In May 2020, the principles and a draft Unicom Pilot Product List were accepted by the members of the Ad Hoc Working Group, allowing the small team to continue to work on this till the end of 2020.

5.3 Selection criteria for the PPL list

The team agreed on the following selection criteria:

- Frequently used active substances with therapeutic intent
- Products from the CEF eHDSI ePrescription Critical Test Data, foreseen to be used in the cross-border ePrescription experiments
- Exemplary complex active substances

The team also decided to exclude:

- Medicinal gasses, radiotherapeutics,
- Active bandages, nutraceuticals, cosmetics or medical devices (e; g; topical products for headlice)

- Medications for veterinary use only
- Magisterial Preparations (compounding by a community pharmacist of a particular pharmaceutical product to fit the unique need of a patient,
- Pharmaceutical Products registered in only a few countries

5.4 Result of the selection: 35 substances

The selection of 35 substances for the UNICOM Pilot Product List (see figure 3) was approved by a wider audience within UNICOM in the summer of 2021.

UNICOM Product Pilot List

1	simvastatin	19	trastuzumab
2	enalapril	20	imatinib
3	omeprazole	21	clomipramine
4	diclofenac	22	carbamazepine
5	cefuroxime	23	metformin
6	salbutamol	24	amlodipine
7	amoxicillin and clavulanic acid	25	perindopril
8	insulin glargine	26	tramadol
9	teriparatide	27	ciclosporine
10	drospirenone and ethinylestradiol	28	itraconazole
11	atorvastatin, amlodipine and perindopril	29	goserelin
12	calcium + vitamin D	30	glyceryl trinitrate
13	paracetamol	31	chloroquine
14	diazepam	32	clotrimazole
15	morphine	33	varenicline
16	enoxaparin	34	ibuprofen
17	hydrocortisone	35	tafluprost
18	lidocaïne		

Figure 3. The UNICOM Pilot Product List: 35 selected substances (incl. 4 combinations)

5.5 Next step: Cleansing of the list

The small team engaged with WP2 (directed by Annet Rozema and team) and dressed a list of all possible modified substances for each of the 35 substances on the list (including information on number of Pharmacovigilance reports for each of the modified substances). EUTCT, UNII, CAS and SNOMED numbers were collected for each modified substances, and the relevant modified substances were listed for each substance.

For each substance, the list of modified substances was critically assessed for relevance, by checking the consistency of being coded across systems, and by looking at the number of pharmacovigilance reports in the EU Pharmacovigilance system.

In a later stage, WHO-UMC added the WHODRUG codes for all the identified substances (either as moiety or modified substance).

The result can be found in **Annex 2,** which lists the following elements:

1. A full result colour-coded worksheet with regard to the data cleansing in WP2, including the comments of the cleansing team and the UNICOM PPL team.
2. An overview of 38 selected moieties (original 35 substances + 3 additional from combination products) coded in EUTCT, UNII, Cas, SNOMED-CT, and INN WHODURG
3. An overview of the modified substances selected, coded in EUTCT, UNII, CAS, SNOMED-CT and INNM WHODURG.
4. A list of INN names with the corresponding WHODURG Code.

What is key in this work is that ***within each coding system the correct code for moiety and modified substance has been identified***. Substances with no modifier were coded at the level of the moiety.

The concept of the "Precise Active Ingredient" is still a subject of debate: its scope definitely encloses all modified substances but possibly also the substances with no modifier.

A discussion is going on about the coding system for the grouper of substances (the INN name). One possible candidate is the WHODURG code for the INN name. The ATC code would not be a sufficient solution for combination products. This issue will be debated at the transatlantic level.

5.6 Gathering medicinal products from 4 countries (NL, BE, UK, GE)

Each member of the small team has listed the available medicinal products packages from their respective medicinal Product Dictionaries for all selected substances, to provide background information on the variability of products in the different markets Only the list of Belgium from the SAM database is publicly available and free of proprietary constraints (see Annex 1).

Together, these lists represent several hundreds of medicinal products and several thousands of medicinal product packages, indicating the already quite substantial coverage of the UNICOM Pilot Product List.

These lists will be used as training sets documenting the full implementation of the IDMP standards at the national level.

5.7 Preparing further work with the UNICOM Pilot Product List

Current discussions are under way to build a FHIR Server, accessible to the agencies, participating in UNICOM, where first results of implementation efforts, based on the PPL can be placed. In work package 9 an analysis was made of the most important descriptive elements from the IDMP framework, to be implemented in national databases, as replacement of addition of existing variables.

In parallel, a thorough analysis of the standardization of the 3 basic elements of IDMP which form the basis for PhPID production had to be performed; it had become clear from the initial discussions in the PPL group that numerous aspects of uncertainty remained insufficiently addressed. In the following chapter, the results of this analysis are presented.

6 Analysis of the 3 basic elements of IDMP (substance, dose form, strength)

In this chapter, we describe analysis and research performed to clarify remaining uncertainties, surrounding the 3 basic concepts of the IDMP standard

6.1 The hierarchy of substances

6.1.1 Physical reality, abstraction, and levels of hierarchy in the representation of substance

During the development of the Drug Ontology in the OBO FOUNDRY, it was recognised that single molecules, even if they pertain to strong medicines, do not heal. The single molecule may have functional properties (a disposition) for instance to bind to a specific receptor. For therapeutic effect to occur (at least as believed in allopathic medicine), a huge number of these molecules in a single pill is needed to expect some therapeutic effect. (Hanna et al., 2016). While each individual molecule has a molecular mass, all the molecules together in a pill (named here somewhat peculiar “a scattered aggregate”) carry a weight, which is a key element of the expression for the strength of the pill. As an example, the active substance of amlodipine is a white crystalline powder, which can be processed into a pill. The weight of the active substance might only be a fraction of the total weight of a pill (a few grams). (See figure 4)

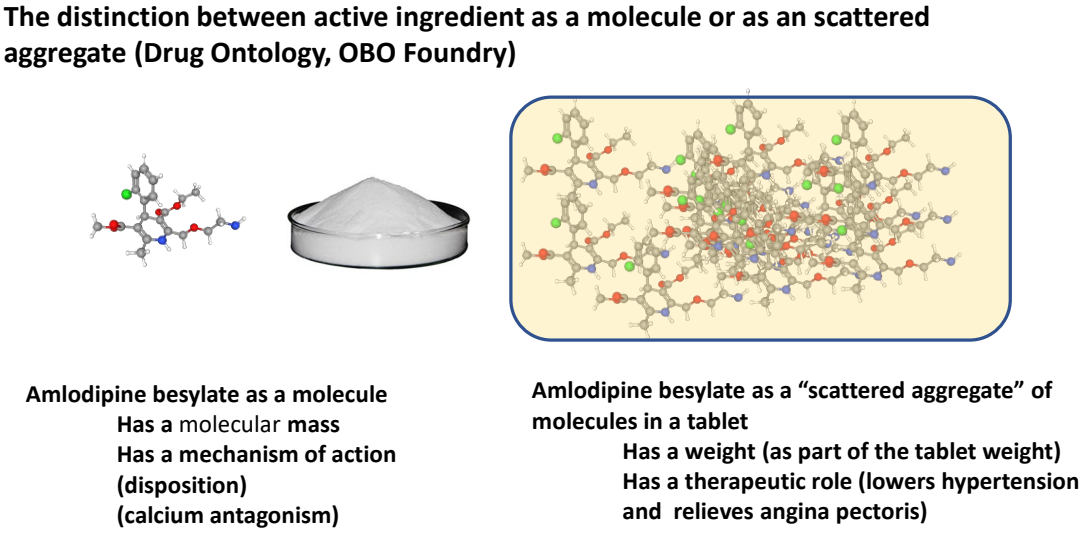


Figure 4. Substance as a molecule and substance as a "scattered aggregate"

In the previous chapters we have learned that amlodipine has 3 different modifiers, amlodipine besylate, amlodipine mesylate, and amlodipine maleate. These substances exist under the form of a powder, which is constituted of two parts: first the moiety (the active part of the molecule) and second the salt. Together they form a physical reality as a white crystalline powder, and can be compounded in solid tablets, and will dissolve in enteric fluids, when ingested. After ingestion and dissolving, the molecule will split in the active moiety (the base) and the salt anion. The active part will then be absorbed, flow into the blood stream, and reach the intended sites of the body.

In Table 8, we listed the structural formulas and the molecular mass of the moiety and of the 3 modified substances, to illustrate the differences, which justifies that in the Chemical Abstract System, each of these physical entities receives a different code.

Table 8. Description of moiety and modified substances of amlodipine

Description of modified susbstances and moiety of amlodipine			
	Formula	Molecular mass	CAS-number
amolodipine	$C_{20}H_{25}Cl_1N_2O_5$	408.9 g/mol	88150-42-9
amlodipine besylate	$C_{20}H_{25}ClN_2O_5 \cdot C_6H_5SO_3H$	567.0 g/mol	111470-99-6
amlodipine mesylate	$C_{20}H_{25}Cl_1N_2O_5 . C_1H_4SO_3H$	505.0 g/mol	246852-12-0
amlodipine maleate	$C_{20}H_{25}ClN_2O_5 . C_4H_4O_4$	524.9 g/mol	88150-47-4

As can be seen from the figures the molecular mass of the moiety is smaller than the modified substances, and each of the modified substances has a different mass.

When looking in fig. 5 at the chemical structure of three modified substances, we can see that there is a common element in the structure, the biggest part is the same in the three modified substances. It is the moiety. Each of the modified substances has a second part, called the modifier. The modifiers are different, and because of that the molecular mass of the modified substance will be bigger than the moiety and different for all three modified substances.

> *Note:*
>
> *To determine the strength of a medicinal product, we will need to know what will be the basis of that calculation. What will be the Basis of Strength Substance (BoSS)? Will it be the moiety or will it by the modified substance. If it is the moiety, products with different modifiers can still be equalised in strength (amlodipine 5 mg per tablet). If the modified substance is the Basis of Strength Substance then there can be subtle differences in the expression of strength of different modified substances for the same moiety. An example is the substance perindopril, marketed by different companies in tablets of 4 or 5 mg, while there is no real difference in the mass of the active moiety, without the salt. This distinction is often misunderstood and can be and was the reason for a lot of confusion in marketing authorisation.*

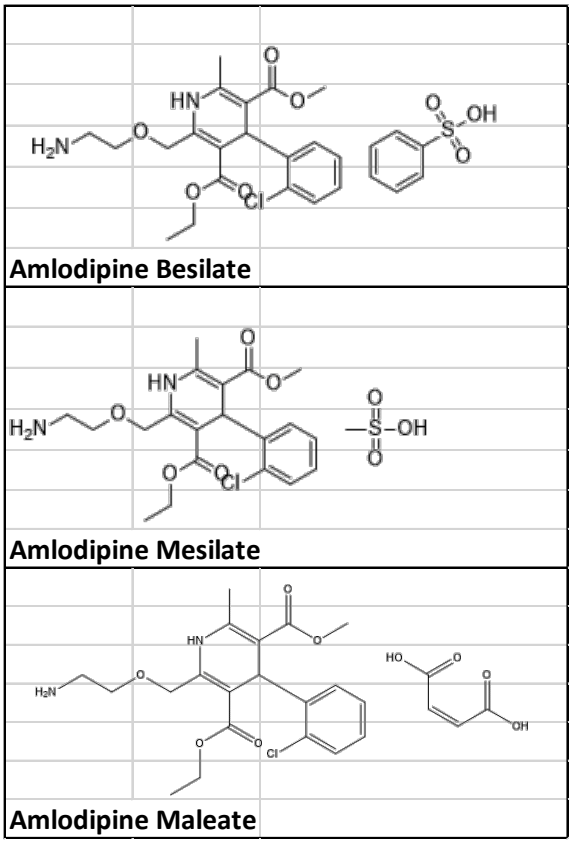


Figure 5. Molecular structures of amlodipine moiety and modifiers

6.1.2 Levels of hierarchy of substances

Having made these fundamental observations, we can now turn to the meaning of the terms that are used to describe different concepts related to substance. Again, we will look at the example of the term “amlodipine”, providing our definition for the different meanings of the same term.

Three meanings of a substance term		
	Amlodipine (1)	
		Term for the physical reality of chemical molecule, which consitutes the active part of an ingredient with therapeutic role. This molecule has a chemical structure, molecular mass, a code in the CAS-system, and a mechanism of action.
	Amlodipine (2)	
		Term for the collection of modified substances (amlodipine besilate, mesilate and maleate), which all contain amlodipine (1)
	Amlodipine (3)	
		Term for the collection of medicinal products that contain any one of the 3 modified substances (named with amlodipine (2)), and no other ingredients with an active role. A medicinal product can be entered in the collection even is the modifier is unknown.
Two meanings of a modified substance term		
	Amlodipine besylate (1)	
		Term for the physical reality of a chemical molecule, consisting of the active part and the salt. This molecule has a chemical structure, molecular mass, a code in the CAS-system, and a mechanism of action
	Amlodipine besylate (2)	
		Term for the collection of medicinal products containing this specific modified substance

Figure 6. Ambiguity in the terms "substance" and "modified substance"

6.1.3 The case of moieties without a modifier

Most of the active substances are chemicals, and most chemicals consist of a moiety and a salt or an ester. Some chemicals do not have a modifier. An example is carbamazepine.

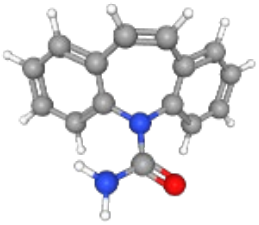


Figure 7. Carbamazepine, a molecule without a modifier

This molecule will exist on its own, will not split in two parts when dissolved, and has only one CAS number (no CAS-number for modified substances). It is a moiety, but not in the sense of the moiety of amlodipine. It is this concept that is needed to accurately describe the substance. If it needs to be presented at the level of a grouper concept, it will be by a collection with only one item.

This situation results in a problem, an unresolved ambiguity in the representation of substances, that is recognised within EMA, FDA, WHO_UMC, and Standard Developing Organisations (SDOs) and hopefully adequately dealt with.

The problem can be best illustrated with an image (see Fig. 10). In case of modified substance it must be determined what the Basis of Strength Substance (BoSS) is: the moiety or the modified substance. In case of a moiety without modifier there can be no confusion.

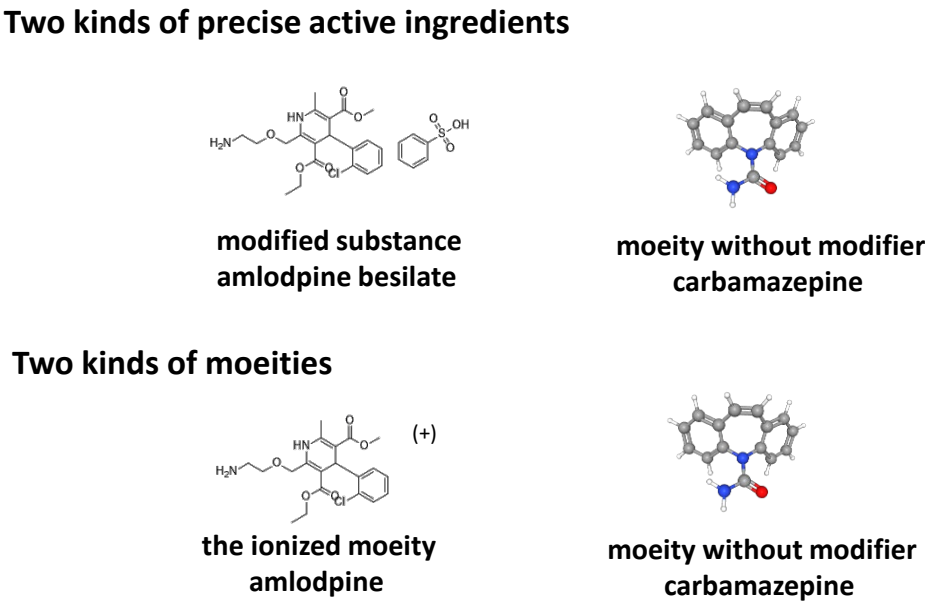


Figure 8. Moiety and Precise Active Ingredient: ambiguity of concepts or of words?

The intense cooperation between experts in WP2, WP8, WP9 of the UNICOM Pilot Product List Project and the international cooperation with WHO-UMC and FDA contributed to raise awareness on these issues. There is a willingness to come to a consensus, and to provide the substance registers of the USA and EU with a solid framework to address the problem of the hierarchy of substances.

6.2 Dose forms and their characteristics

6.2.1 Existing coding systems for dose forms

There are 3 coding systems for standardizing dose forms of medicinal products: Snomed-CT, RxNorm, and EDQM (advocated by the IDMP standards).

The dose form model of Snomed-CT is similar to EDQM, with terms, definitions, characteristics, and an ontological structure. The granularity is slightly less (which may be beneficial), the value sets for the characteristics are slightly different (which is remediable), and the definitions are more consistently formalized.

In RxNorm, the dose form is actually a value set to describe characteristics of products. It is much less granular (179 dose forms in RxNORM, vs 428 in EDQM). Definitions are not very precise. There is a rudimentary collection of dose form types, which tend to overlap.

Note: WHODrug uses New Form Code of the European Pharmaceutical Marketing Research Association, another standardised dose form coding system. Also CDISK has controlled terminology for dose forms.

6.2.2 The value of EDQM

EDQM Standard terms for Pharmacology originates from the European Pharmacopoeia, with a long tradition of excellence in maintaining the controlled vocabularies.

It is a terminology for Europe, but it has been recognised internationally by the International Committee for Harmonisation (a platform for the EU, US, and Japan) for global standardization in pharmaceutical issues of regulation and research.

The ISO standard regarding dose forms is currently under revision in ISO WG6 of TC125, coinciding with the analyses performed in UNICOM.

For the global identification of medical products, ISO has opted for a granular description of the dose form, and the granularity of EDQM (428 human terms) is much deeper than the terms list of RxNorm (179 terms).

The value of the standardized use of a core set of characteristics for the dose form (needing transformation, release characteristics, intended site, and administration method) is widely recognised also by the FDA, and ICH.

A pilot project has been conducted by the FDA to use the characteristics of the EDQM Dose Forms, as the basis for PhPID production, and this has been explored in collaboration with WHO-UMC (See **Annex 3**)

Hence the time is right for a serious reflexion on how this excellent European resource can grow into a global terminology.

6.2.3 Possible improvements for EDQM in the transition to a global terminology

Within WP8 in UNICOM (IDMP and Clinical Care), an analysis was performed of EDQM, in preparation of pilots to use IDMP in decision support systems.

This analysis was initiated by Robert Vander Stichele, I-HD, lead of WP8, together with a semantic expert (Joseph Roumier) and a drug database expert (Dirk Vannimwegen). It is an integral part of this deliverable, and can be consulted in **Annex 4**.

Here we will limit ourselves to a brief summary.

First, a selection was made of the terms for human use, leading to an analysis of 428 current dose forms.

The following minimal interventions were explored:

- All pharmaceutical dose forms that need transformations were explicitly aligned with their resulting administrable dose form (and basic dose form and state of matter). This work was particularly important as the IDMP standard stipulates explicitly that for the calculation of the ***PHPID-L4, the code of the administrable dose form must be used***, ***and that the strength must be expressed as the strength of the administrable dose form***.
- All definitions were analysed and compared with a set of the characteristics for definitional completeness and consistency
- Value sets of the characteristics were analysed for multiplicity (more than one value for a characteristic of a dose form). A new value was created combining the values of the constituent values (instead of storing the possible values in separate columns).
- The dose forms with the value “cutaneous/transdermal for intended site were critically analysed, and the replacement with a single value (either cutaneous or transdermal), was explored which was feasible in almost all the cases.
- The intended site of sublingual dose forms was oro-mucosal, while there is a clear intention to reach the systemic circulation with these dose forms (sometimes intentionally sublingual to bypass the first-liver pass). The value “sublingual” was suggested.
- An exploration was made concerning the possibility to also characterise the dose forms as having a systemic or local effect. A few dose forms were identified where that was problematic, in the cutaneous/transdermal, the nasal, and the rectal dose forms, but for the vast majority of dose forms this was perfectly possible.

We came to the conclusion that a thorough analysis was needed of the definitional value of the characteristics and its use in the PhPID production process (as proposed by the FDA).

To be able to perform this analysis, an experimental new version of the EDQM database of terms was created, implementing the changes described above. This database and the implicit structure were transferred to WebProtegé, an online ontology manager, also used during the transition development from ICD10 to ICD11.[5] This allowed ontological work by collaborating experts, using the groupware facilities of WebProtégé. In addition, the data were also transferred in an SQL database, for further analysis.

The 428 lines of the EDQM Human Dose Form Database were then ordered on the four characteristics (transformation, Release Characteristics, Intended Site, Administration Method).

Collections of dose forms with the same four values for their characteristics were constructed. These collections were then analysed separately for all collections with the same intended site, and by needing transformation or not. The result was manually analysed and meaningful groups were formed or split. Basic criteria for group formation were the following ones: ***similarity of characteristics, clinical relevance, and similarity of the way strength would be expressed*** for such a group. This also resulted in further suggestions for improvement of the basic data, but also in an initiation of the steps to develop an ontology of dose forms.

[5] https://protegewiki.stanford.edu/wiki/WebProtege

6.2.4 Possibility of creating a dose form ontology.

The 428 dose forms of EDQM can already be divided into groups sorting them on the 25 values of the intended site (oral, nasal, rectal, etc). We proposed and developed one intermediate level of granularity, depending on the similarity of the characteristics as described above.

The result is described in the **Annex 4**, and will be discussed further with experts of EDQM and formally presented to WG6 of ISO/CEN.

6.2.5 Comparison to RxNorm and Snomed-CT dose forms

For the comparison with RxNorm we refer to the study report in **Annex 5.** It is obvious that the dose terms of RxNorm are used as descriptive variables, and do not constitute an ontology on their own. The level of granularity is much less (only 144 dose forms), and the definitions are broad. There is no formal system of characteristics, but the FDA is contemplating to use the EDQM characteristics (see WHO_UMC/FDA Pilot Report in Annex 3). There is a rudimentary but overlapping list of drug form groups.

Preliminary discussions were initiated with the Dose Form experts of SNOMED-CT, mainly to explain the approach, and exchange ideas on mapping or alignment between the 2 systems.

6.2.6 Application to the UNICOM Pilot Product List

As is opted for the granular EDQM dose form codes, ***any dose form encountered during analysis of national medical products can now be standardized to an EDQM dose form term***, with the corresponding codes and characteristics. The question remains whether in the SPOR data register the code for dose forms will be the existing EDQM code, or a new proprietary code from EMA.

6.3 The complexity of determining strength of medicinal products

Historically, in the regulatory process there are many examples of inconsistencies and discrepancies in the determination of the strength of similar medicinal products, and when considering the emergency authorisation of the first Covid-19 vaccines, this proved to be agin the case..

In the report of the WHO_UMC/FDA report (see Annex 3) also inspired by the EU IDMP Implementation guide[6], many challenges have been listed with remedial suggestions.

The following tables and figures of this paragraph were extracted from the WHO_UMC/FDA Pilot report to demonstrate the approach to find a practical way to implement the complex framework developed in IDMP, to be able to represent it in a standardized way all the possible variations of strength expression.

6.3.1 Strength determination and substance

First, 3 different kinds of strength are defined: presentation strength, concentration strength, and reference strength (see Table 9).

Table 9. Strength definitions

Strength definitions	
Strength (Presentation)	When the strength of a substance is described as a qualitative term describing the discrete unit in which a Pharmaceutical Product is presented
Strength (Concentration)	When the strength of a substance is expressed as the amount of substance per unit of measurement, such as millilitre or gram
Reference Strength	The strength for the active moiety. If there is no reference substance, the active substance and its strength must not be repeated in the data object reference strength

It is clear that correct and standardized determination of strength depends predominantly on a thorough insight in the molecular mass of moieties and modified substances, and for that concrete guidance is given in the report with examples (See table 10).

[6] https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/products-management-services-implementation-international-organization-standardization-iso-standards_en.pdf

Table 10. Calculation of the reference strength

If in the SmPC the active substances are given as salts/esters or pro-drugs and the strength corresponds to the salts/esters or pro-drugs,	the reference substance strength is being calculated as follows: ○ Substance (mg)/(substance molecular weight(mg/mol)* reference substance molecular weight(mg/mol))
If in the SmPC the active substances are given as salts/esters or pro-drugs and the strength corresponds to the active moiety,	the salts/esters or pro-drugs substance strength is being calculated as follows: ○ Reference substance (mg)/(reference substance molecular weight(mg/mol)* substance molecular weight(mg/mol))
If in the SmPC the active substance is the active moiety.	In this case the active substance is identical to reference substance and no calculation is performed.

This can again be illustrated with the example of amlodipine. (See table 11)

Each of the 3 modified substances (salts) of amlodipine have a different molecular mass. But in the end the authorized strength that will appear on the package is 5mg.

It is the reference strength, in this case weight of the moiety of amlodipine in one tablet.

In the production process of pills, each company using a particular modified substance will have to calculate (using the formula's described above) ***how much of the modified substance needs to be sprinkled into a tablet***, to reach the exact reference strength of 5 mg. In table below, the result of this calculation is given up to 4 decimals of the weight (as this has to be an exact calculation for a precise production process).

Table 11. Calculation of the weight of different modified substances for a given reference strength for the moiety amlodipine

	Reference strength		Molecular mass of the moiety		Molecular Mass of the Modified Substance		Weight of the scattered aggregate of the modified substance in the tablet	
Amlodipine besylate	5	mg	409	g/mol	567	g/mol	6,9315	mg
Amlodipine mesylate	5	mg	409	g/mol	505	g/mol	6,1736	mg
Amlodopine maleate	5	mg	409	g/mol	530	g/mol	6,4792	mg

6.3.2 Strength and dose form

The way strength is expressed depends also on the type of dose forms. Oral drops and oral solutions, are both liquids, but their strength will be expressed differently, as will the strength of skin creams and dermal patches. The FDA and WHO_UMC have in their pilot explored these issues, and proposed already five different patterns of dose forms, relevant for the expression of strength (see figure 9). This work is a pilot, to be discussed at the international level, also with EMA, and with the European agencies. For UNICOM it is an opportunity to test these proposals with medicinal products of the Product Pilot List, and to provide feedback in the consensus seeking process for this delicate endeavour.

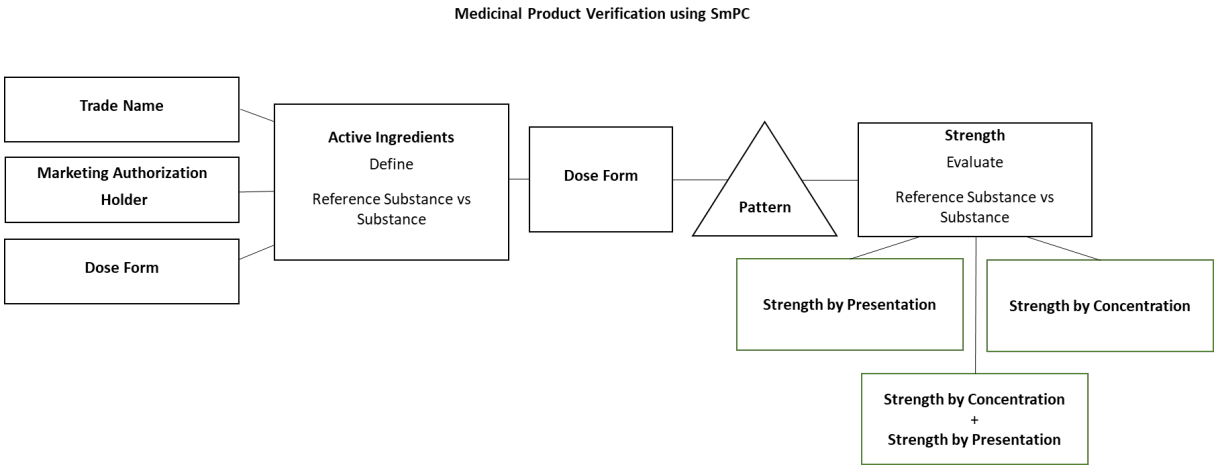


Figure 9. Determining strength based on business rules (pattern) dependent on type of dose form and nature of substance.

In the WHO_UMC/FDA pilot and the EU Implementation Guide a number of these patterns have already been developed (see table 12)

Table 12. Concatenated Patterns Framework

	Type of product	Examples	Pharm Prod. unit of Present	Strength by Presentation	Strength by Concentration
A	Solid, countable Solid dose forms in "container" Metered dose delivered by a metered actuation	Tablets, capsules, suppositories Powder or granules in sachet, Inhalers, Spray	Tablet, capsule Container (sachet, etc.) Actuation (inhaler, etc.)	Mandatory	Empty

B	Unit dose or continuous presentation (dosing is individual/not accurate and the total volume in the container is of less importance for dosing purposes)	Vials, Unit dose solutions, parenteral liquid, unit dose nebuliser solutions	N/A since it is the concentration that is relevant	Empty*	Mandatory
C	Products enclosed in a "presentation", where the dose has a delivery rate	Transdermal patches	Patch	Empty	Mandatory– as a delivery rate over time
D	Unit dose or continuous presentation, diluted to different final concentration depending on application such as Injection and/or infusions	Vials, parenteral liquid	Container (vial, etc.)	Mandatory	Mandatory

*Note that in order to simplify for PhPID generation ***the approach is to express the strength as Strength per Concentration only***, ignoring expressing strength according to unit dose.

A further investigation is needed to make sure that the appropriate number of patterns of dose forms is determined. It is important to form groups of dose forms with similar characteristics and the same business rules to determine strength. The ontological analysis of dose forms of EDQM, as described supra, might be instrumental to this.

The UNICOM Pilot Product List will provide a good sample of medicinal products to experiment with these new insights.

6.4 Application to the UNICOM Pilot Product List

An effort has been undertaken to start to build a collection of triplets of precise active ingredients, granular dose form, and strength, looking at the available medicinal products in the 4 test countries.

By analysing the Belgian products pertaining to 3 of the 35 PPL substances, we analysed for each substance which (modified) substance were marketed, which dose forms, and which strengths. Codes from WHO_UMC and EUTCT were added to the modified substances and moieties with no modifier. Dose Forms were standardized for EDQM and the terms and codes of the administrable dose form and pharmaceutical dose form were added. Structured strength descriptions were entered.

For the identified triplets, the VTMGroups (available in the Belgian Database) and ATC codes were gathered, together with the WHODRUG Code for the INNname.

In Table 13, the result of this latter step is described, with an overview of the available modified substances and substances without modifier.

Table 13. Identification of available modified substances, dose form, strength and corresponding high level concepts and codes of international classification ATC.

Information gathered for Belgian Medicinal Products in preparation for basic IDMP implementation (substance, dose form, strenght) for amlopine, amoxicilline, carbamazepine							
Elements needed for higher levels of abstraction							
VTMGROUP	**WHODrug**	**ATC**	**ATC_ROA**	**VMPGROUP**	**Substance**	**modified**	**WHO_UMC**
amlodipine	00972401001	C08CA01		amlodipine oral 10 mg	amlodipine	besilate	'00972401001
					amlodipine	besilaat	'00972401001
					amlodipine	maleate	00972403001
					amlodipine	maleate	00972403001
				amlodipine oral 5 mg	amlodipine	besilate	'00972401001
					amlodipine	besilate	'00972401001
amoxicilline	00249601145	J01CA04	P	amoxicilline injection/infusion 1 g	amoxicilline	natrium	00249603001
		J01CA04	O	amoxicilline oral 125 mg / 5 ml	amoxicilline	trihydrate	00249602001
				amoxicilline oral 1 g	amoxicilline	trihydrate	00249602001
					amoxicilline	trihydrate	00249602001
				amoxicilline oral 250 mg / 5 ml	amoxicilline	trihydrate	00249602001
				amoxicilline oral 500 mg	amoxicilline	trihydrate	00249602001
					amoxicilline	trihydrate	00249602001
				amoxicilline oral 500 mg / 5 ml	amoxicilline	trihydrate	00249602001
carbamazepine	00052501001	N03AF01		carbamazepine oral 100 mg / 5 m	carbamazepine		'00052501001
					carbamazepine		00052501001
				carbamazepine oral 200 mg	carbamazepine		'00052501001
				carbamazepine oral (gereg. afgift	carbamazepine		'00052501001

In the ATC classification, for some codes distinctions are made between the oral or parenteral use of the substance (mostly because the calculation of the DDD is different in those two instances.

When this was the case, this was indicated, and aligned with the appropriate dose form(s) of the substance. Also, it is to be foreseen that for some substances, different ATC codes will need to be identified, depending on the dose form (and exceptionally on strength). According to ATC framework, this is only the case when there are clearly different therapeutic uses.

As in Belgium, the operationalisation of INN Prescribing is officially regulated, and integrated in the National Medicinal Product Dictionaries SAM and the BCFI database, this item was easily retrievable.

In table 14, we gather all the basic elements that could be fed into a procedure for PhPID production, and aligned with the correct codings of WHODRUG INNM and EUTCT for substance, and EDQM for Dose Form, together with the structured strength descriptions.

Table 14. Identification of available triplets with codes of their 3 constituents (substance, dose form, strength)

Elements needed for preparing the procedure of PhPID production											
Substance	modified	WHO_UMC	Substance_ code EUTCT	ADF_ code EDQM	EDQM ADF	PDF_code	PDF_ term	str_Value	Str_unit	Str_ deno_ val	Str_ denom _unit
amlodipine	besilate	'00972401001	100000090079	10210000	Capsule, hard	10210000	Capsule, hard	10	mg		
amlodipine	besilaat	'00972401001	100000090079	10219000	Tablet	10219000	Tablet	10	mg		
amlodipine	maleate	00972403001	100000089370	10219000	Tablet	10219000	Tablet	10	mg		
amlodipine	maleate	00972403001	100000089370	10220000	Coated tablet	10220000	Coated tablet	10	mg		
amlodipine	besilate	'00972401001	100000090079	10210000	Capsule, hard	10210000	Capsule, hard	5	mg		
amlodipine	besilate	'00972401001	100000090079	10219000	Tablet	10219000	Tablet	5	mg		
amoxicilline	natrium	00249603001	100000090113	50060000	Solution for injection/infusion	50053500	Powder for solution for injection/infusion	1	g	2,50000	mL
amoxicilline	trihydrate	00249602001	100000092629	10106000	oral suspension	10111000	Powder for oral suspension	125	mg	5	mL
amoxicilline	trihydrate	00249602001	100000092629	10104000	oral liquid	10121000	Dispersible tablet	1	g	50	mL
amoxicilline	trihydrate	00249602001	100000092629	10221000	Film-coated tablet	10221000	Film-coated tablet	1	g		
amoxicilline	trihydrate	00249602001	100000092629	10105000	oral suspension	10111000	Powder for oral suspension	250	mg	5	mL
amoxicilline	trihydrate	00249602001	100000092629	10210000	Capsule, hard	10210000	Capsule, hard	500	mg		
amoxicilline	trihydrate	00249602001	100000092629	10104000	oral liquid	10121000	Dispersible tablet	500	mg	50	mL
amoxicilline	trihydrate	00249602001	100000092629	10105000	oral suspension	10111000	Powder for oral suspension	500	mg	5	mL
carbamazepine		'00052501001	100000092127	10106000	Oral suspension	10106000	Oral suspension	100	mg	5	mL
carbamazepine		00052501001	100000092127	10117000	Syrup	10117000	Syrup	100	mg	5	mL
carbamazepine		'00052501001	100000092127	10219000	Tablet	10219000	Tablet	200	mg		
carbamazepine		P	100000092127	10226000	Prolonged-release tablet	10226000	Prolonged-release tablet	200	mg		

This information can now be fed into the first experimental pilots for actually producing via a Hash function the first actual PhPIDs.

7 The procedure for production of PhPIDs

Figure 10. Haiku on what binds and seperates almost similar things

7.1 The need for global governance

Some have advocated that the rules for producing global identifiers for medicinal products, based on 3 basic elements (substance, dose form, strength) would be so simple and self-explanatory, that every agency of the world would be able to operate this on their own and still produce an identical result as in everywhere else in the world.

More likely, is the ***proposal to dedicate the task of the production of PhPID to a robust supra-national organisation***, ***while the standardized basic data should be collected in cooperation with the agencies.***

But before deciding on which organisations may operate the procedures, it must be decided how the procedure itself should run.

7.2 What has already been done

WHO-UMC, who was initially not a formal partner in the UNICOM project, was invited by WP1 (the coordination of the work of SDOs in UNICOM) to participate in internal discussions and in transcontinental exchange of ideas. WHO-UMC and FDA together engaged in a pilot project, initiated by the will of the FDA to work on the dose forms, but resulting in experiments of producing PhPIDs, already in emergency times for COVID-19 vaccines, and for some example drugs. For a full version of this report see Annex 3.

7.3 Pathway to a consensus on the procedure for PhPID production

The aim of UNICOM is to have full sets of PhPIDs, at least for the medicinal products of the UNICOM Pilot Product List, and preferably for most of the countries, participating in UNICOM. And this by the time the pilot projects for cross-border ePrescription will be rolled out.

The EU implementation Guide is currently not informing about what to do with the PHPID, and it may well be that decisions will not be reached before the end of the UNICOM project.

Hence, it would be wise to accept an experiment and initiate pilots where initial ideas can be tested and inform the process of writing the implementation guide.

WHO_UMC and FDA have paved the way with their pilot.

UNICOM (WP1) and WHO-UMC have written a draft procedure for the production of PhPIDs, for a period of step-wise testing. First, the specifications for this procedure need to be completed, the draft proposal adapted to a first round of analysis, and then tested on the UNICOM Pilot Product List. The experience gained, and the invested made, should be corroborated in a sustainability plan, to make sure that this endeavour is continued, also after the UNICOM project has ended. ***WHO-UMC, a robust organisation, with a long track in drug terminology, and a vital role in pharmacovigilance, has expressed interest to take responsibility in this issue***, in transatlantic cooperation, in cooperation with UNICOM and ISO, during the coming pilot years, and thereafter. In **Annex 6**, a first attempt at clarifying the specifications for the draft production of PHPIDs is given: it is a living document where the results of the WHO_UMC/FDA still have to be incorporated.

While the first proposed PhPIDs start to roll out, it would be advisable to keep track of the result of that production process so that if can be shared other, and be used by the persons working on concrete applications using the UNICOM Pilot Product List. In the next chapter, we describe a UNICOM tool whose objective is precisely to make this possible.

8 Building a UNICOM Repository of validated Pharmaceutical Product Identifiers - PhPIDs

8.1 Rationale

The result of the UNICOM Pilot Project in terms of draft identifiers (with preliminary validation) will need to be collected.

WP9 has proposed a FHIR resource for assisting agencies in experiments to implement full implementation of IDMP in their databases, by means of test FHIR server, with procedures based on compliance with the EU Implementation Guide, now in V2, and soon to be updated.

However, it is worthwhile to organize a limited data collection of the basic elements which are needed for PhPID production. In the Pilot Product project, triplets of basic data have already been collected for medicinal products of 3 substances on the list, to be submitted to a draft production of PhPID by WHO-UMC, during the UNICOM project. The resulting PhPIDs can be stored in a public repository (preferably in LOD (Linked Open Data), containing:

- The draft PhPID identifiers
- The codes and terms of their 3 basic constituents
- The substance hierarchy
- The dose form ontology
- The links to pharmacotherapeutic classifications

8.2 ICT-Operationalisation

A small team in WP8 (Task T8.1) worked out the ICT- model for this repository, with a conceptual model (see figure11), a logical model (see figure 12), and a technical model

8.2.1 Conceptual model

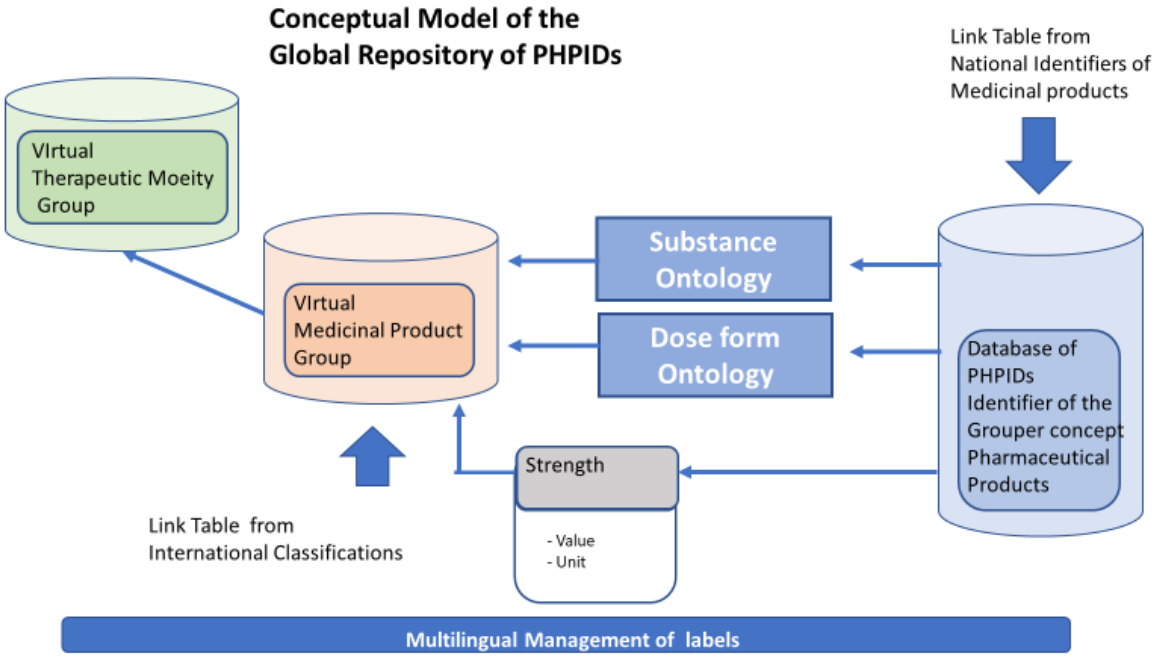


Figure 11. Conceptual model of a repository of PhPIDs

In this conceptual model, the database of draft PhPIDs is the kernel of this system. It contains the PhPIDs and their constituencies (modified substances, dose forms, strength). National drug dictionaries can provide a link to this database for those products where there is a PhPID available, and provided actual medicinal products for this PhPID are on the national market.

There is a possibility to group similar PhPIDs to a higher level of abstraction, called Virtual Medicinal Product Group (e.g. amlodipine oral 5 mg). This abstraction is based on very simple ontologies. One aggregation is proposed for substance to make the aggregation of modified substances into a class of substances, containing the same moiety. Another simple ontology is for dose form, based on the characteristic for Intented Site, but with an intermediate level grouping to the granular EDQM dose form (see Annex 4).

The link with international classifications will be established through this Virtual Medicinal Product Group, which will greatly reduce the workload of constructing the chain from high level international classifications to actual medicinal products in a country.

The model also foresees a Virtual Therapeutic Moiety concept, which is an abstraction of the PhPID_level 1 (only substance).

In these models no abstractions are foreseen for PhPID-Level 2 (substance + strength, considered as not really relevant), and for PhPID_Level 3 (Substance + dose form).

However, concepts similar to these dose forms in Snomed-CT and RxNorm are easily programmable with this approach.

The model foresees multilingual management of the value sets of the crucial variables, such as substance, dose form, units of measurement, of course, using as much as possible the EUTCT, EDQM, and SPOR facilities.

8.2.2 Logical model

The logical model deals with:

- The handling of combination products (in the PhPID concepts, Virtual medicinal Product Group Concept, and the Virtual Therapeutic Moiety Concept)
- The precise implementation of the substance and dose form ontology.

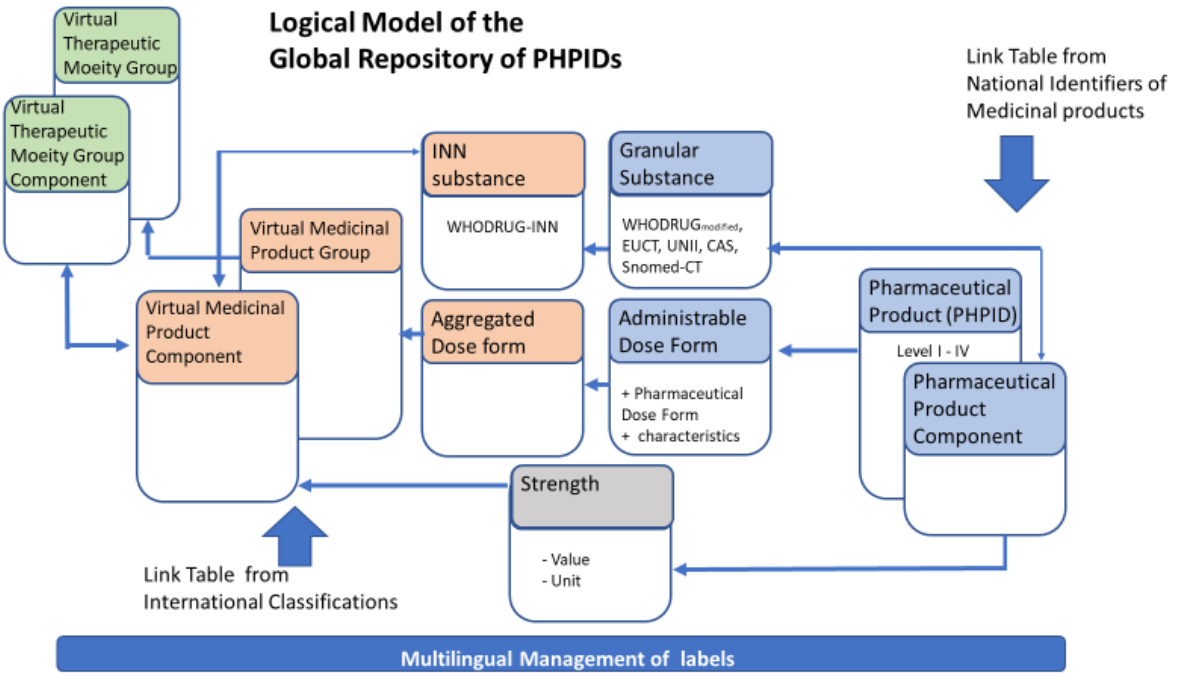


Figure 12. Conceptual model of a repository of PhPIDs

8.2.3 Technical model

The full technical model of this database has been elaborated, in the aftermath of this deliverable, and will be populated with the results of the first experiments of the PhPID production.

The repository can be populated with data from medicinal products, pertaining to the substances on the PPList, originating from a limited number of countries, to start with.

8.3 Fostering the impact of UNICOM on clinical care and research

The ultimate goal of this approach is the very objective of this deliverable: **finding a robust interoperable way of connecting precise medicinal product identification in the different constituencies with the international classifications.**

Each country in Europe and also elsewhere in the world has in most cases several databases with their medicinal products. The Marketing Authorisation Agency usually has a database collecting data on the process of accepting New Drug Applications and changes to the labelling (variations). Countries may have an official Drug Database, governing the processes of eHealth. In some countries, there is an independent drug information Centre, which produces web sites and printed material for health professionals and patients. Pharmacist associations, vendors of Electronic Health Records may all have their own drug database. International and national publishers of scientific information may also provide drug dictionary services.

Some international publishers coordinate the maintenance of the connections to a number of national drug databases in an internal proprietary system, which is a costly endeavour.

Developers of Decision Support Systems rely on international classifications to build their rules for decision support. However, connecting these rules to different national drug dictionaries is not an easy task, which hampers the diffusion of these valuable resources through the European market. If UNICOM can make a contribution here, the efforts will have been worth it.

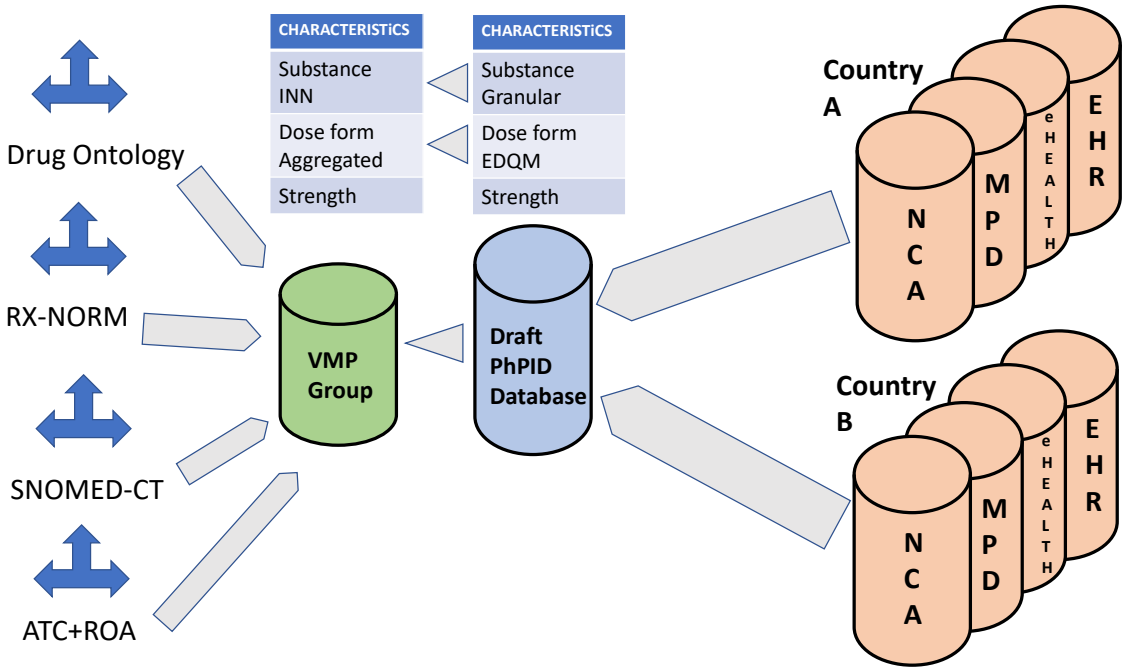


Figure 13. Connecting national drug dictionaries with international classifications

8.4 Prospects on sustainability of this UNICOM repository, beyond completion

Such an accessible, open repository can provide valuable examples of linking of national medicinal products to existing PhPIDs and to international classifications. As the content grows, it can become more and more interesting for agencies and publishers to consult. It would also be instrumental to the pilot project of WHO_UMC and FDA, to publish their results. In a spirit of transferability, this repository can be planned to be handed over to supranational organisations, by the end of UNICOM.

9 Linking drug classifications to IDMP

9.1 A short and incomplete inventory of international drug classifications

Pharmacotherapeutic classifications play a very important role in the organisation of drug information, labelling, pharmacovigilance, clinical care and research. Pharmacopoeia and handbooks of pharmacology may have complicated tables of content, while the Anatomical Chemical Therapeutic (ATC) Classification of the World Health Organisation (WHO) is a straightforward taxonomy with only 5 levels[7].

While clinical classification reflects the medical culture and therapeutic thinking, the ATC provides a tool for correct recording of the consumption of medicinal products (grouped in broad classes) during specific time periods in specific regions.

At the world level, for pharmacovigilance applications, the Uppsala Monitoring Centre of the World Health Organisation maintains the WHODRUG dictionary, including Standardised Drug Groups (SDG)[8] :

Drug Classifications linked to national medicinal product dictionaries exist in many countries. In the USA, RxNorm is maintained by the National Library of Medicine, and integrated with the Medical Subject Headings (and its pharmaceutical classes) and UMLS (Unified Medical Language System)[9].

First Databank (FDB Med knowledge is a proprietary medicinal product dictionary, integrated with decision support systems for the USA, with international ambitions[10] . Other example are the British National Formulary[11], Vidal in France (partner in UNICOM)[12], Rote Liste in Germany[13] , the Belgian Centre for Pharmacotherapeutic Information[14], and in the Netherlands Z-index[15] and Farmacotherapeutic Kompas[16], a compilation of labelling information at the level of the INN substance.

Snomed-CT has an important drug class component and its own model to represent medicinal products with the possibility to link to national medicinal product dictionaries through national extensions. It provides the key to integrating drug information into medical documentation[17].

The Drug Ontology DrOn, related to the OBO Foundry of biomedical ontologies, is an embryonic attempt to integrate semantic web ontologies with description of medicinal products and drug classification[18]. In the Linked Open Data Network, the information from many pharmaceutical and chemical databases are intricately linked[19]. Creating reliable links between international drug classifications and national medicinal products

Several international attempts exist to link the information resources mentioned above with a broad range of national medicinal product dictionaries.

On the semantic web, there is a Linked Open Data Network LOD , linking together all semantic web databases, including life sciences[20].

[7] https://www.whocc.no/
[8] https://www.who-umc.org/whodrug/whodrug-portfolio/whodrug-global/
[9] https://www.nlm.nih.gov/research/umls/rxnorm/overview.htm
[10] https://www.fdbhealth.com/solutions/medknowledge-drug-database
[11] https://www.bnf.org/
[12] https://www.vidal.fr/
[13] https://www.rote-liste.de/
[14] https://www.cbip.be
[1515] https://www.z-index.nl/g-standaard
[16] https://www.farmacotherapeutischkompas.nl/
[17] https://confluence.ihtsdotools.org/display/DOCMPM),
[18] http://www.obofoundry.org/ontology/dron.htm
[19] https://lod-cloud.net/
[20] https://lod-cloud.net/#life_sciences

Numerous national drug databases for regulation, science and clinical practice have links to the ATC classification, as it is the standard for producing drug statistics for international comparison. This (almost) ubiquitous link will probably be very important to build the bridge between national medicinal products and the proposed repository of PhPIDs. In addition, accurate description of pack size will help national actors to calculate more precisely the number of Defined Daily dose per package. This is a crucial operation to ensure standardised international comparison of drug consumption[21] .

The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) is based on a conformational process of national drug data to the common standard of RxNorm. It is widely used in international pharmaco-epidemiological research[22].

The promise of IDMP implementation is that it will bring ease and precision to the process of linking national medicinal products and product packages to all the International classifications.

9.2 The need for a simple classification of indications and products for Medical Education and Patient Information

Medical students, pharmacy students and nursing students must acquire a minimal working knowledge of the therapeutic arsenal, during their graduate and post-graduate training. Universities have selected a basic set of ailments, with which their students must become familiar, and also selected the pharmacotherapeutic classes, suitable for more intense teaching and for testing the achievement of educational objectives.

The Netherlands and Belgium share an electronic platform for case-based, problem-oriented teaching of pharmacotherapeutics. At the kernel of this education instrument is a compact database of indications (the 300 most important in primary care, organised in 2 levels) and pharmaceutical classes (350 in also organised in two levels), linked to each other, so that choice of a particular indication will evoke a list of potentially pertinent pharmaceutical classes and substances. Currently this information resource is linked to the 5th Level of the ATC classification (the level of substance). In the second half of the UNICOM project, it is planned to link also to PhPIDs, to facilitate the international implementation of this electronic educational platform, in cooperation with the European ENLIGHT program[23].

Finally, for patients, names of pharmacotherapeutic classes may be an enigma. And yet, drug labelling contains many cross-references to pharmacotherapeutic classes. For instance, the labelling of amlodipine mentions that patients who are allergic to dihydropyridines should not take amlodipine. But how can the patient actually know whether one of the medications on his/her allergy list is a dihydropyridine?

Building a controlled vocabulary of drug group names, with professional terms and laymen equivalents (either laymen terms or descriptions), has been realised in an old European Project, called Multilingual Medical Glossary, still available on the Web[24].)

Combined with the indication/pharmaceutical class resource described above, and linked to PhPIDs, this could provide the basis for international apps that allow patients to navigate drug labelling, and to interprete their medication list.

These resources (currently in development) will be made available through the software factory, organised by WP6 in Unicom.

[21] https://www.whocc.no/use_of_atc_ddd/

[22] https://www.ohdsi.org/data-standardization/the-common-data-model/

[23] https://enlight-eu.org/

[24] https://users.ugent.be/~rvdstich/eugloss/welcome.html

10 Linking decision rules in pharmacotherapy to drug classifications and IDMP

10.1 Explicit rules for (in)appropriate prescribing from internationally validated PIM-lists.

In the past decades tools have been developed to assist researchers and health care professionals in analysing the quality of their prescribing. With the greying of society and the extension of the preventive and therapeutic arsenal, polypharmacy has become more and more prevalent in the population in general, among the older adults, and especially among the oldest old. More people reach advanced age, are confronted with multi-morbidity, and also with polypharmacy. For prescribers, finding the balance between appropriate treatment of multiple ailments and avoidance of the downside of polypharmacy is a delicate exercise.(Hoffmann et al., 2020)

Several international validated tools for assessing the quality of prescribing have been developed, either as survey tool for field research, either as ICT-applications for clinical care. The latter are software tools that will perform a pre-analysis of medication lists of patients, resulting in a number of comments to the prescribing physician who will then decide to follow the advice or not.

Explicit criteria from 3 internationally validated lists of explicit criteria for potentially (in)appropriate medication Lists were collected and brought together in a web-based repository. (Ivanova et al., 2018)

- the EU-7 PIM list (geared to medications on the European markets;
- the Beers' list (developed in the US and used widely in the world),
- the STOPP/START list (developed by Irish geriatricians, also looking at underuse).

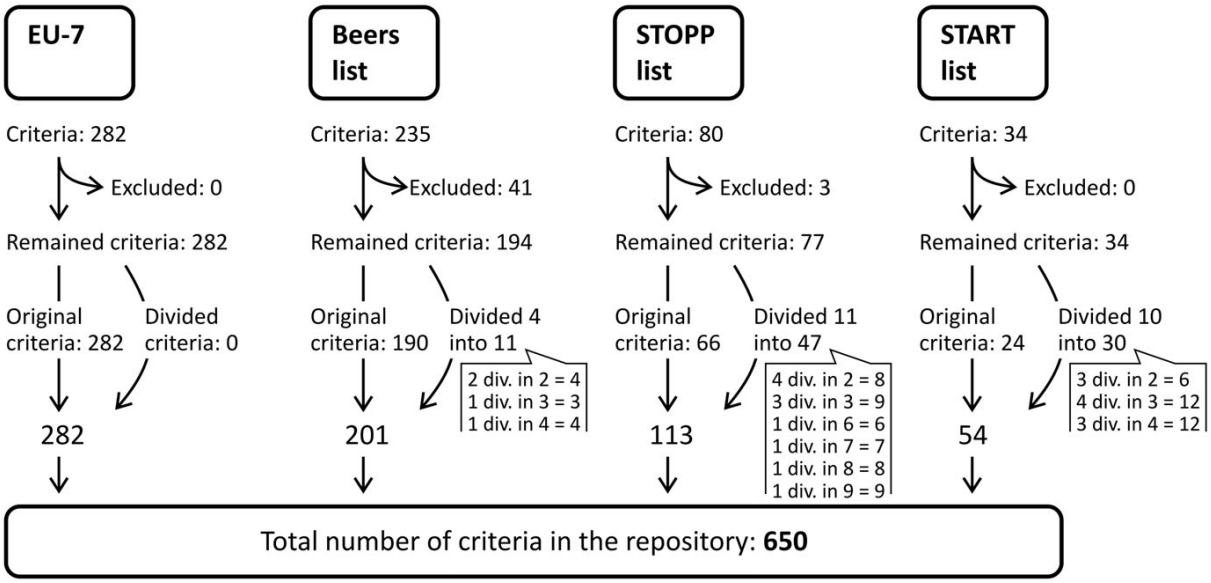


Figure 14. Selection process of criteria in a repository from 3 validated lists of explicit criteria of potentially inappropriate prescribing

10.2 Sources of clinical data and medications to apply the criteria

Looking at the literature of studies, having used these tools, the main terminology systems were International Classification of Diseases and the ATC for medications, with the algorithms working on the medication list and on the problem list of the Electronic Health Record. ***As most criteria are rather crude, the information captured in an International Patient Summary could be sufficient for operating these criteria in a valid way***.

10.3 Coding the rules to disease and medication identifiers

The criteria are in the following format: IF the patient has disease X THEN Drug Y should not be taken BECAUSE Drug Y is contra-indicated for Disease X.

Each mention of medication in the criteria in the repository has been recoded to the pertinent list of ATC codes. Hence, these links could be extended to IDMP; using the repository of PhPIDs, as proposed above.

For the disease element in the rules, several classification systems (ICPC, ICD, SNOMED-CT) can and should be used, given the variety of implementation in clinical systems.

Criteria with medications pertaining to the UNICOM Pilot Product List will be expressed in collections of PhPIDs and this information will also be made available in the software factory of WP6 by the second half of the UNICOM project. **Knowledge drug databases *will have to consider IDMP as a pivot between decision rules and medicinal product dictionaries*** (Hoffmann et al., 2020) (Hoffmann et al., 2020)(Eiermann et al., 2010) .

10.4 Piloting IDMP decision support in UNICOM

It should be possible in the second half of the project to select a number of explicit PIM criteria, pertaining to medication on the UNICOM Pilot Product List, for experiments with International Patient Summaries, to be demonstrated in HL7 Connectathons.

10.5 Using explicit criteria to test the quality of clinical data in the International Patient Summary (iPS)

In many health institutions (such as nursing homes), the medication list is amongst the most reliable data on the patients (often the only reliable data, available in code form). The quality of the clinical date (in the list of problems, diagnoses, indications) is often more problematic.

The element of medication in the explicit criteria could be used to scan the medication list for relevant medicines. This can trigger requests for relevant clinical information (e.g. what is the indication for this medication? Is a well-known contra-indication for this medication present in the problem list?)

These requests can then help the health care provider to complete the information in International Patient Summary, to make sure that it becomes a document that better reflects the current condition of the patient, and is a better basis for fair medical audit of the prescribing quality.

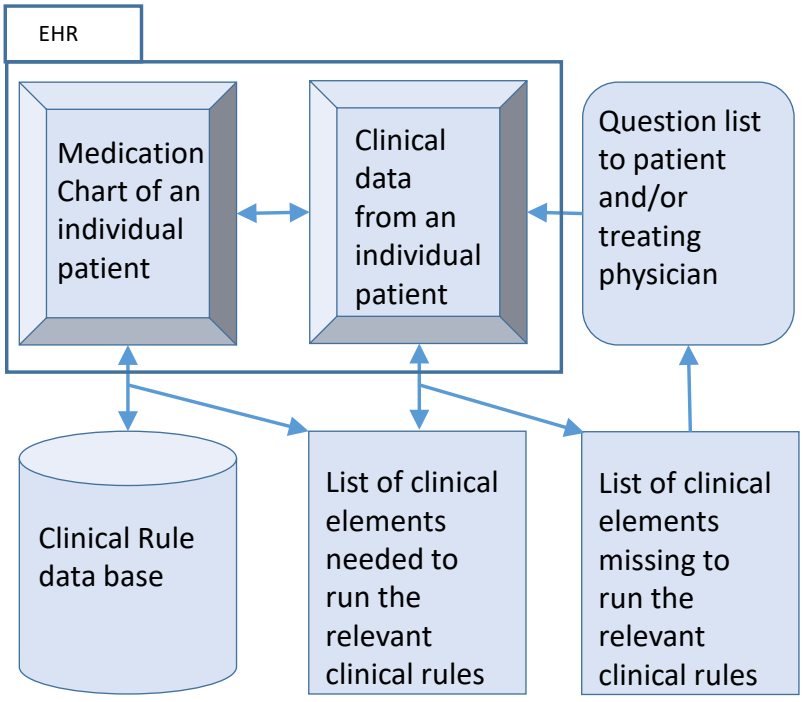


Figure 15. Audit of quality of clinical documentation with clinical rules for appropriate prescribing

11 Conclusions

The objective of UNICOM is the implementation of the ISO/CEN standards across the continuum of the Research and Development departments of the industry, the marketing authorisation eHealth drug Databases, the Medicinal Product Dictionaries, and the databases of vendors of Electronic Health Records and International Patient Summaries, and finally the data warehouses of the health care sector.

The precise and standardised description of many essential details of each individual medicinal product in a given constituency is a crucial prerequisite.

Three elements are of fundamental importance: substance, dose form, and strength. UNICOM will be instrumental in the pilots where coding systems will be chosen, decisions on granularity of information will be made, and procedures developed to feed HASH-functions to produce global identifiers, called Pharmaceutical Product Identifiers PhPID level 1 to 4.

Full implementation of IDMP for the detailed description of all aspects of a medicinal product requires a full toolbox to service the requirements of specific use cases.

A PhPID_Level IV (combining substance, dose form and strength) will groups medicinal products from different constituencies. The PhPID aspect of UNICOM can be the Swiss knife approach within IDMP, to allow application of the standards across different domains and use cases, while assuring interoperability between domains and systems.

In this deliverable and in the Gap Analysis of WP1, a number of unresolved issues have been identified that stand in the way of operation routing production of global identifiers, such as the hierarchy of substance, the granularity of dose forms, the representation of strength according to substance definition and dose form pattern, and the handling of combination products.

Methodological approaches to solve the remaining issues have been suggested, while stressing the need for international cooperation, involving also the national marketing authorisation agencies.

Next steps are to insure demonstration of draft procedures for the production of the PhPID, based on examples from the UNICOM Pilot Product List.

Building and maintaining a trustworthy repository of Pharmaceutical Product Identifiers will be crucial to initiate the applications that will demonstrate the value of IDMP.

Figure 16. Pharmaceutical Product Identifier (PHPID-Level IV) as the connecting element between domains and use cases (courtesy of Christian Hay, UNICOM, WP1).

12 References

Aronson, J. K., & Aronson, J. (2012). *Balanced prescribing-principles and challenges Correspondence*. https://doi.org/10.1111/j.1365-2125.2012.04413.x

de Vries, T. P. G. M., Henning, R. H., Hogerzei, l H. V., & Fresle, D. A. (1995). *Guide to Good Prescribing*.

Denig, P. (1994). *Drug choice in medical practice* [University of Groningen]. https://research.rug.nl/en/publications/drug-choice-in-medical-practice-rationales-routines-and-remedies

Denig, P., & Haaijer-Ruskamp, F. M. (1992). *Therapeutic decision making of physicians* (pp. 14:9-15). https://doi.org/10.1007/BF01989219

Dhombres, F., & Bodenreider, O. (2016). *Interoperability between phenotypes in research and healthcare terminologies-Investigating partial mappings between HPO and SNOMED CT*. https://doi.org/10.1186/s13326-016-0047-3

Eiermann, B., Bastholm Rahmner, P., Korkmas, S., Landberg, C., Lilja, B., Shemeikka, T Veg, A., Wettermark, B., & Gustafsson, L. (2010). Knowledge Bases for Clinical Decision Support in Drug Prescribing – Development, Quality Assurance, Management, Integration, Implementation and Evaluation of Clinical Value. In *Decision Support Systems*. InTech. https://doi.org/10.5772/39458

Hanna, J., Bian, J., & Hogan, W. R. (2016). An accurate and precise representation of drug ingredients. *Journal of Biomedical Semantics*, 7–7. https://doi.org/10.1186/s13326-016-0048-2

Henning, R. H., Hogerzeil, H.V., Kosabasoglu, Y. E., Smith, A. J., Summers, R. &, & De Vries, T. P. G. M. (2001). *Teacher's Guide to Good Prescribing.* https://apps.who.int/iris/handle/10665/67010

Hoffmann, M., Vander Stichele, R., Bates, D. W., Björklund, J., Alexander, S., Andersson, M. L., Auraaen, A., Bennie, M., Dahl, M.-L., Eiermann, B., Hackl, W., Hammar, T., Hjemdahl, P., Koch, S., Kunnamo, I., Le Louët, H., Panagiotis, P., Rägo, L., Spedding, M., … Gustafsson, L. L. (2020). Guiding principles for the use of knowledge bases and real-world data in clinical decision support systems: report by an international expert workshop at Karolinska Institutet. *Expert Review of Clinical Pharmacology*, 1–10. https://doi.org/10.1080/17512433.2020.1805314

Ivanova, I., Elseviers, M., Wauters, M., Christiaens, T., & Vander Stichele, R. (2018). European repository of explicit criteria of potentially inappropriate medications in old age. *Geriatrics and Gerontology International*, *18*(8), 1293–1297. https://doi.org/10.1111/ggi.13331

Maxwell A, S. R. (2016). Rational prescribing: the principles of drug selection. *CME CLINICAL PHARMACOLOGY*, *16*, 459–464.

Senger, C., Seidling, H. M., Quinzler, R., Leser, U., & Haefeli, W. E. (2011). Design and evaluation of an ontology-based drug application database. *Methods of Information in Medicine*, *50*(3), 273–284. https://doi.org/10.3414/ME10-01-0013

Tichelaar, J., Richir, M. C., Garner, S., Hogerzeil, & H., & De Vries, T. P. G. M. (2020). WHO guide to good prescribing is 25 years old: quo vadis? *Eur J Clin Pharmacol*, *76*, 507–513. https://doi.org/10.1007/s00228-019-02823-w

Van Bever, E., Wirtz, V. J., Azermai, M., De Loof, G., Christiaens, T., Nicolas, L., Van Bortel, L., & Vander Stichele, R. (2014). Operational rules for the implementation of INN prescribing. *International Journal of Medical Informatics*, *83*(1), 47—56. https://doi.org/10.1016/j.ijmedinf.2013.09.004

Van Doorn, A., RBJ, Brouwers, R., Fulto, A., & De Graef, P. (2009). A National multicentre e-learning programme to promote cooperation and to promote rational prescribing by future physicians and pharmacists. *Basic and Clinical Pharmacology and Toxicology*, *105*, 40–40.

Vander Stichele, R. H., Elseviers, M. M., Ferech, M., Blot, S., Goossens, H., & ESAC Project Group. (2004). European surveillance of antimicrobial consumption (ESAC): data collection performance and methodological approach. *British Journal of Clinical Pharmacology*, *58*(4), 419—428. https://doi.org/10.1111/j.1365-2125.2004.02164.x

Winnenburg, R., & Bodenreider, O. (2014). *A framework for assessing the consistency of drug classes across sources* (Vol. 5). https://doi.org/10.1186/2041-1480-5-30

Wirtz, V. J., Mol, P. G. M., Verdijk, J., Vander Stichele, R. H., & Taxis, K. (2013). Use of antibacterial fixed-dose combinations in the private sector in eight Latin American Countries between 1999 and 2009. *Tropical Medicine & International Health : TM & IH*, *18*(4), 416—425. https://doi.org/10.1111/tmi.12068

13 List of Annexes

13.1 D8.1_Annex 1. List of Medicinal products from BE for the 35 substances from the UNICOM Pilot Product List

Is an Excel spreadsheet, as attachment in Zip file

13.2 D8.1_Annex 2. Result of data cleansing of substances on the Unicom Pilot Product List

Is an Excel Spreadsheet, as attachment in Zip file

13.3 D8.1_Annex 3. Report of the WHO-UMC/FDA Pilot for PhPID production

13.4 D8.1_Annex 4. Analysis of EDQM terminology

13.5 D8.1_Annex 5. Comparison EDQM / RxNorm

13.6 D8.1_Annex 6. Specifications for the Draft Procedure for PhPID production

Document version	D8.1_ANNEX 3_ Report of the WHO-UMC/FDA Pilot for PhPID production
Authors	Malin Fladvad, UMC, Ron Fitzmartin, U.S. FDA Ta-Jen Chen, U.S. FD Terry Quinn, NCI / EVS Julia Nyman, UMC
Changes	Referred to with permission in the context of the UNICOM Project

U.S. FDA / UMC Pilot to Evaluate the Use of Pharmaceutical Dose Form Characteristics for Pharmaceutical Product Identification

OUTCOME REPORT

Authors: Malin Fladvad, UMC, Ron Fitzmartin, U.S. FDA, Ta-Jen Chen, U.S. FDA, Terry Quinn, NCI / EVS, Julia Nyman, UMC

Version 2.0

Date : May 20, 2021

Version History

Preliminary Note	April 23, 2021	
Procedures Version 1.1	April 29, 2021	Malin Fladvad Julia Nyman
Procedures Version 1.2	June 09, 2021	Malin Fladvad Julia Nyman

Content

Introduction

The ISO standards for Identification of Medicinal Products (IDMP) describe the use of standardised definitions for the identification and description of medicinal products, such as, identification of Substances, Medicinal Product (MPID), Medicinal Product Package (PCID) and Pharmaceutical Product (PhPID) aiming to facilitate the exchange of medicinal product information in a robust and consistent manner. Implementation of the standards on a global level would improve interoperability across global regulatory and healthcare communities, ensuring unambiguous communication across jurisdictions.

To have consistent identification according to IDMP, the substance ID and PhPID should be maintained at a global level. The Uppsala Monitoring Center (UMC) has, after discussions with WHO and regulators, agreed to generate and maintain global PhPIDs for pharmacovigilance processes as well as other applicable use cases. The PhPID is generated based on information about substance, administrable dose form and strength, all defined in the ISO IDMP standards. The pilot project will focus on an evaluation of predefined global dose form characteristic codes as input into the generation of PhPID.

The original ISO 11239:2012 standard, Regulated Information on Pharmaceutical Dose Forms, Units of Presentation, Routes of Administration and Packaging, was based on the use of a single controlled vocabulary (i.e., EDQM) that regions could use to communicate with each other. Further, the technical specification, TS 20440:2016 guide for ISO 11239:2012, assumes the use of a single controlled vocabulary or a mapping to it. However, there is currently no agreement on such a central vocabulary, in particular, for pharmaceutical dose forms.

Various regions are using their own set of terminologies for dose form, which are not harmonised and show different levels of granularity between regions, making a one-to-one mapping between a regional terminology and a centrally controlled vocabulary of low quality. Investigations have shown around 20-45 % one-to-one matches between the ISO 11239 compliant EDQM terminology and terminologies used by U.S. FDA (FDA) and Health Canada. Similar results were also shown for mappings to SNOMED, CDISC and EPMRA dose form terminologies.

To solve the issues with mapping between different dose form terminologies, a proposal was made at International Standards organisation TC 215 WG6 October 2020 meeting to use a centrally maintained set of dose form characteristics to describe a dose form term and code, for use in global IDMP, and in generation of PhPID.

To evaluate this new concept for description of dose form, U.S. FDA and UMC agreed to perform a pilot study. The pilot would:

1. Assign chosen EDQM dose form characteristics for US marketed medicinal products corresponding to their selected substances,
    a. Substances identified in the UNICOM Pilot product list[1]
2. Evaluate the performance of the dose form characteristics for generation of corresponding PhPID
    a. Generation will use numerical representations of dosage form together with substance and strength.

[1] The UNICOM pilot product list has been created within the UNICOM project, a European Commission supported Innovation Action focused on the implementation of the ISO IDMP standards. The list contains medicinal substances that represent the range of challenges that exist for medicinal product identification information and the objectives for the list is to provide actual exemplar content for cross border perscription pilot as well as other areas of work within UNICOM.

Objective

The objective of the pilot is to demonstrate that a selected set of dose form characteristics from EDQM and other potential characteristics can describe dose forms and be utilized as input in the generation of PhPID.

Project Scope

According to the ISO standard for Pharmaceutical product, ISO 11616, PhPID shall be presented for both active substance and specified substance, each containing four PhPID identification levels. The PhPID shall be generated using the corresponding ISO standards and technical specifications:

Substance ISO 11238 and ISO/TS 20440

Administrable dose form ISO 11239

Units of measure ISO 11240

Table 1. Four levels of PhPID

Table 1 — Four levels of PhPID

PhPID active substance stratum	PhPID_SUB_L1 → Substance(s) PhPID_SUB_L2 → Substance(s) + strength + reference strength PhPID_SUB_L3 → Substance(s) + administrable dose form PhPID_SUB_L4 → Substance(s) + strength + reference strength + administrable dose form
PhPID specified substance stratum	PhPID_SpSUB_L1 → Specified substance(s) PhPID_SpSUB_L2 → Specified substance(s) + strength + reference strength PhPID_SpSUB_L3 → Specified substance(s) + administrable dose form PhPID_SpSUB_L4 → Specified substance(s) + strength + reference strength + administrable dose form

This project focuses mainly on exploring PhPID for active substance on the fourth level, calculating PhPID_SUB_L4 according to table 1.

The data on medicinal products included in the pilot has been provided by U.S.FDA based on the substance data from UNICOM Pilot product list.

The dose forms used for the PHPID calculation are expressed according to four of the centralized core EDQM dose form characteristics (and their codes); Release characteristics, Intended Site, and Administration Method and Basic administrable dose form.

The process for generation of PHPID includes the following steps:

1. Assignment by the FDA of the centralized core dose form characteristics for US marketed medicinal products corresponding to the substances in the UNICOM Pilot product list.
2. The UMC validate the FDA data and assign the relevant strength to generate corresponding PhPID, using numerical representations of dose form characteristic together with substance and strength;
3. FDA and UMC will perform a data equivalency assessment on the use of centralized core dose form characteristics and other potential characteristics for generation of PhPID.

Data Processing Procedures

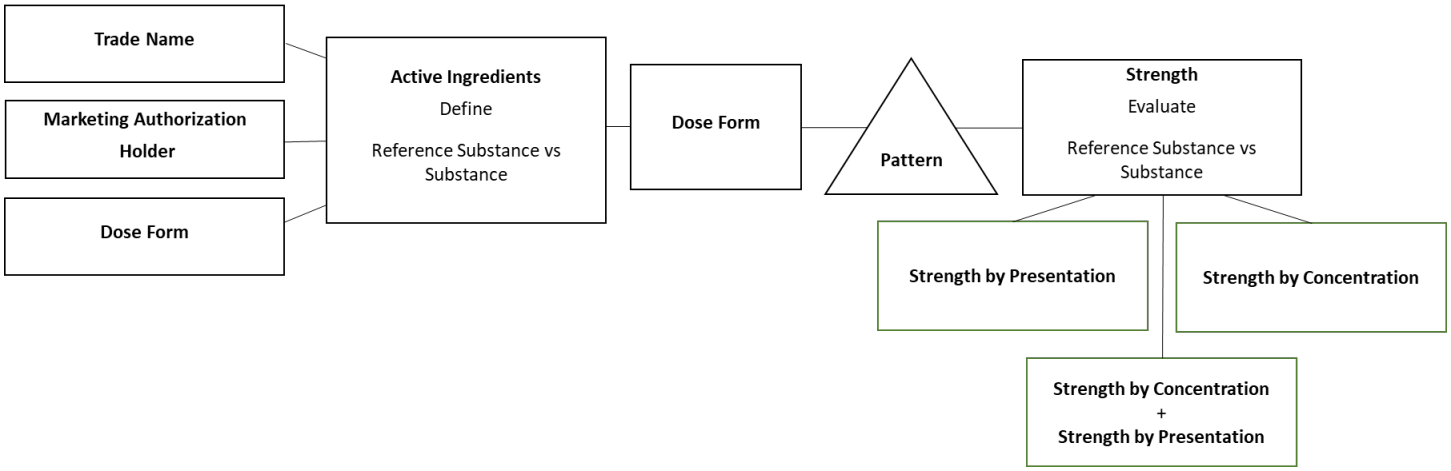
The purpose of this section is to describe the process applied to collect and validate the data to be used for PhPIDs construction.

The data, provided by U.S. FDA, included 1167 US marketed medicinal products with assigned centralized core dose form characteristics (Release Characteristics, Intended Site, Administration Method and Basic Administrable Dose Form) and information on active ingredient, trade name, Marketing Authorization Holder, route, state of matter, transformation information, UNII code, approved strength.

In order to generate necessary input data for PhPIDs construction, information on substance/reference substance, dose form, unit of presentation, unit of concentration, and strength has been collected. In order to validate the data, the following process has been applied:

1. Substance information has been verified in SmPC, selected based on Trade Name, Marketing Authorization Holder, and Dose Form information, and relevant salts/esters of the substance have been defined appropriately;
2. Dose form information has been evaluated further in order to select a suitable pattern, using Pattern Framework.
3. The strength information has been evaluated depending on pattern selected in the step above - in some cases the strength (presentation) or strength (concentration) or both have been entered;
4. Strength information has been further evaluated in the SmPC, to determine if the strength corresponds to the active substance or to the reference substance (active moiety); If the strength corresponds to the active substance, the strength value has been filled in the substance field and reference substance strength (active moiety strength) value has been calculated in the field for reference strength and vice versa.

Figure 1. Medicinal product verification process using SmPC.



Combination products (combination of substances)

In case of combination products and how their respective substances and strength will be presented, each combination product has been assigned a unique ID number to identify the input data for all substances and their respective strength in the combination.

Substance

This section provides and overview on the procedures applied to analyze information on active substance, assumptions made, and challenges faced during the validation process.

The first task in the validation process was to identify the active substance with therapeutic intent at the most appropriate level of granularity.

For chemicals this means that the active substance must be defined, for example, as an active moiety or a salt/ester. If the active substance is a salt/ester or similar, a reference substance needs to be identified, i.e. active moiety. For biological substances and vaccines, a simplification was made where the substance was identified as the active moiety and no reference substance was considered.

For identification of the substance to be submitted to the HASH function in this pilot, the WHODrug ID was used. For future use of a global substance ID, the level of granularity needed for unique identification of a substance will be based on the current investigation of minimal fields by ISO WG6.

There were several challenges regarding substances (Table 2) that have been faced and corresponding assumptions taken for this pilot to mitigate the risks of PhPID generation inconsistency.

Table 2. Challenges and their description vs mitigation related to substance evaluation process

Challenge Description	Challenge Mitigation
In case substances have been inadequately described in the label	If a substance is inadequately described in the label, the active moiety is chosen as an active substance. Examples: 1. Ibuprofen and ibuprofen potassium, Trade name=Ibuprofen from SOFGEN PHARMACEUTICALS LLC and HUMANWELL PURACAP PHARMACEUTICAL WUHAN CO LTD. The substance was presented as a mixture between the free acid and salt, but no ratio was given.
Hydrates	When there is any inconsistency in the substance description, if it is a hydrate or not, a pharmacopeial monograph or a structure connected to an official name such as INN or USAN can be used as a reference. If there is no clear guidance, the non-hydrate form is used. The active moiety is used as the reference substance in either case. 1. Morphine sulfate – according to both USAN and Ph Eur the name Morphine sulfate is the pentahydrate, hence the pentahydrate is used in this pilot. 2. Amoxicillin – according to USAN the name Amoxicillin is the trihydrate and the Ph Eur only have a monograph for the trihydrate, hence the trihydrate is used in this pilot. 3. Amoxicillin (unspecified), the non-hydrate was chosen 4. Omeprazole Magnesium –according to USAN and Ph Eur there is only the non-hydrate variant and is hence used instead of the trihydrate mentioned in the label text. 5. Lidocaine HCl – the non-hydrate was used unless a clear description in the label (if investigated) stated otherwise see example Zingo https://www.accessdata.fda.gov/drugsatfda_docs/label/2007/022114lbl.pdf According to USAN the name Lidocaine HCl is the monohydrate and the Ph Eur has Lidocaine HCl Monohydrate, but since there were labels as in the example above the name sent by FDA was used. 6. Estradiol, Trade name= ANGELIQ. The substance was described as the non-hydrate in the text, as well as the name and molecular weight. But the structure was represented as the hemihydrate.
INN named salts	If a salt is assigned an INN name, like Enoxaparin sodium then the salt is considered the active moiety and no reference substance was used for the PHPID calculation.

Covid vaccines approach	For the COVID-19 vaccines included in the dataset, the substance was identified as the active moiety and no reference substance was used.
Similar biological substances approach	It has been decided to consider these substances as having the same substance identifier. Further specifications are possible according to more detailed substance information, but have not been taken into account for the PHPID production within pilot´s scope (i.e. substances defined as identical, for all trastuzumab containing products).
Product name vs active ingredients data source FDA	Only substances included in the active ingredient list are considered as a basis for the PHPID. Example: The product name included both Lidocaine and Dextrose (i.e. 5% Lidocaine Hydrochloride and 7.5% Dextrose Injection), but only lidocaine was included as an active ingredient in the label.

Table 3: An example on differences between pharmacopeias and naming bodies regarding the naming of hydrates.

Substance	**Pharmacopeia/naming body**	**Name**	**Structure and additional information**
Lidocaine hydrochloride, CAS 73-78-9	Ph Eur	NA	
	USAN	Lidocaine hydrochloride, anhydrous	As part of the (USAN name) CAS 6108-05-0 record
	JAN	Lidocaine hydrochloride	
	KP	NA	
Lidocaine hydrochloride monohydrate, CAS 6108-05-0	Ph Eur	Lidocaine hydrochloride monohydrate	
	USAN	Lidocaine hydrochloride	
	JAN	NA	
	KP	Lidocaine hydrochloride hydrate	

There are several similar examples for other hydrates where the pharmacopeias and naming bodies don't align. There are also inconsistencies regarding naming within the different pharmacopeias.

Dose Form

Core Characteristics

This section provides an overview on how centralized core EDQM dose form characteristics (Release Characteristics, Intended Site, Administration Method and Basic Administrable Dose Form and their codes) and other potential characteristics can be used to describe dose forms for use in the generation of PhPID, assumption made, and challenges faced.

There are several challenges regarding dose form (please see table 4) that have been faced and corresponding assumptions taken for this pilot in order to mitigate the risks of PhPID generation inconsistency.

Table 4. Challenges and their description vs mitigation related to dose form evaluation process

Challenge Description	Challenge Mitigation
Dose Form expression variations	When dose forms are expressed differently within different jurisdictions as seen for some of the Covid-19 vaccines, a clear base for decision of which dose form should be used to create a global PHPID is necessary. Please find example of dose forms assigned for the Covid-19 vaccine Comirnaty within different authorities: **Authority of approval** — **Administrable dose form** EMA — dispersion for injection FDA — suspension for injection UK — solution for injection
Label information variations	When medicinal product dose form description is twofold in SmPC, in case the product consists of a capsule that should be opened and swallowed and is described as granules - the dose form is treated as a capsule similar to the concept for a solid dose form in a container (i.e. Alkindi sprinkle)
Label information variations	When medicinal product dose form description is twofold in SmPC, in case the product is described as a system and as a patch – the patch has been selected as a dose form (i.e. Ztlido)
Unit of Presentation	For intradermal injection system select "system" as unit of presentation according to EDQM definitions (i.e. Zingo)

Pattern Selection	Pattern A has been selected for nasal sprays that deliver its entire contents upon activation (i.e. Valtoco)
Pattern selection	Pattern A has been selected for enema products, where applicator should be considered as UOP (i.e. Cortenema)

To summarize, a framework for handling both dose form variations between different regions as well as label variations needs to be developed for consistent PHPID generation.

Strength

This section provides guidance on how to record and/or express information on strength, and reference strength of active ingredients present in medicinal products, assumptions made, and challenges faced.

As outlined in ISO 11616, in order to unambiguously link the strength to the product - both strengths for the substance and reference substance (when the active ingredient is a salt/ester/pro-drug) are deemed. The identification of substance or reference substance and their corresponding strength has been verified in the SmPC. The reference strength is derived from active moieties of an active substances(s).

- If in the SmPC the active substances are given as salts/esters or pro-drugs and the strength corresponds to the salts/esters or pro-drugs, the reference substance strength is being calculated as follows:
    - Substance (mg)/(substance molecular weight(mg/mol)* reference substance molecular weight(mg/mol))
- If in the SmPC the active substances are given as salts/esters or pro-drugs and the strength corresponds to the active moiety, the salts/esters or pro-drugs substance strength is being calculated as follows:
    - Reference substance (mg)/(reference substance molecular weight(mg/mol)* substance molecular weight(mg/mol))
- If in the SmPC the active substance is the active moiety. In this case the active substance is identical to reference substance and no calculation is performed.

Strengths are defined according to ISO 11240 describing the use of the UCUM standard, where possible. When the strength of a medicinal product needs to undergo a transformation (for example reconstitution) for administration, the strength resulting from the transformation accordance with the regulated product information (i.e. in the SmPC) is stated. There are several challenges regarding substances (please see table 4) that have been faced and corresponding assumptions taken for this pilot in order to mitigate the risks of PhPID generation inconsistency.

Table 5. Challenges and their description vs mitigation related to strength evaluation process

Challenge Description	**Challenge Mitigation**
Strength expression variations	In order to create consistent PHPIDs, the strength expression must be harmonized for similar products. When there is any inconsistency in the strength expression for similar products the UCUM standards and/or SI units where selected when possible: Strength given in % (i.e. Diclofenac epolamine) is expressed in mg

	Strength given in IU (i.e. Ergocalciferol) is expressed in mg when possible Strength given as mg/g or mg/mL (i.e. locoid solution) is expressed in mg/g Unit variations for biosimilars (i.e. Hercipin Hylecta) is expressed in IU
Hydrates- Label information variations affecting strength expression	The value of the reference strength may vary depending on if the substance is a hydrate or non-hydrate and this information is often not clear as shown in the examples in table 2. A harmonized expression of hydrates needs to be developed to ensure consistent PhPID assignments
Strength for reference substances	The strength for reference substances has been not expressed when the active substance is an active moiety.
Patches or similar without a delivery rate	Patches and systems that do not include a delivery rate are placed under the pattern A and the strength is expressed as presentation. Other patches/film, extended release that have delivery rate are placed under pattern C. Please find below some examples on strength expression for patches/systems, sourced from FDA data file:

Active Ingredient	**Trade Name**	**Dosage Form**	**Strength**	**Pattern**	**Strength Expression**	**Substance Strength**	**Ref Subs Strength**
Diclofenac epolamine	Flector	Patch	1,3%	A	Strength by Presentation	180mg	130mg
Lidocaine	Lipoderm	Patch	5%	A	Strength by Presentation	700mg	N/A
Lidocaine	Ztlido	Patch	1,8%	A	Strength by Presentation	36mg	N/A
Lidocaine hydrochloride monohydrate	Zingo	System (intradermal)	0,5 mg	A	Strength by Presentation	0,5mg	0,4mg
glyceryl trinitrate	Nitro-Dur	Film, extended release	0.4mg/hr	C	Strength by Concentration	400ug/h	N/A

Strength expression variations	When strength is expressed differently within different jurisdictions as seen for some of the Covid-19 vaccines, a clear base for decision of which strength expression should be used to create a global PHPID is necessary.

	Please find example of expression of strength for Covid-19 vaccine from AstraZeneca within different authorities: Authority of approval — Strength per dose (0.5 ml) EMA — 2.5×10^8 infectious units UK — 5×10^{10} viral particles Australia — 5×10^{10} viral particles
Concentrate´s strength	In cases medicinal product is formulated as a concentrate and shall be diluted with an unknown amount of liquid (i.e. water, juice) for oral administration, it is challenging to obtain an accurate concentration after the dilution. Therefore, concentrate's strength of the concentrate before dilution has been used for PhPID generation (i.e. Diazepam Intensol™ Oral Solution (Concentrate)).
Strength interval	The use cases for expression of strength interval needs to be clarified. Would it for example be applicable to products that are administered according to parameters such as age and weight or products dissolved differently depending on administration (in intramuscular or intravenous)?

Pattern Framework

According to ISO 11616, the strength used for the PhPID is primarily the presentation strength. For liquid preparations, the standard suggests that both the presentation strength (expressed as total volume of the container) and the concentration strength (per unit volume, giving a value of 1 in the denominator) should be taken into account. A separate PhPID called the product code concept is suggested to represent the strength concentration and should be mapped to the strength presentation at all applicable PhPID levels. This pilot has evaluated the above concepts for strength and their extent of integration in the PhPID.

Table. 6 Strength definitions

Strength definitions	
Strength (Presentation)	When the strength of a substance is described as a qualitative term describing the discrete unit in which a Pharmaceutical Product is presented

Strength (Concentration)	When the strength of a substance is expressed as the amount of substance per unit of measurement, such as millilitre or gram
Reference Strength	The strength for the active moiety. If there is no reference substance, the active substance and its strength must not be repeated in the data object reference strength

To determine the nature of the strength expressions and the above-described issues for different types of product, the EU IG described patterns for expressions of Pharmaceutical Product, will be evaluated[2]. The patterns show how the Manufactured Item (MI) and the Pharmaceutical Product (PhP) should be expressed for a certain type of product. Products can then be matched to the appropriate pattern which then demonstrates how the MI and PhP should look like, for which the strength is mandatory.

The data validation process showcased dissimilar practices in different regions in expressing the strength and lead to a need for further simplification of the patterns has been identified and the Concatenated Patterns Framework has been obtained.

The following reference table provides the necessary guidance to select between Presentation strength and Concentration Strength for different types pf products:

Table 7. Concatenated Patterns Framework

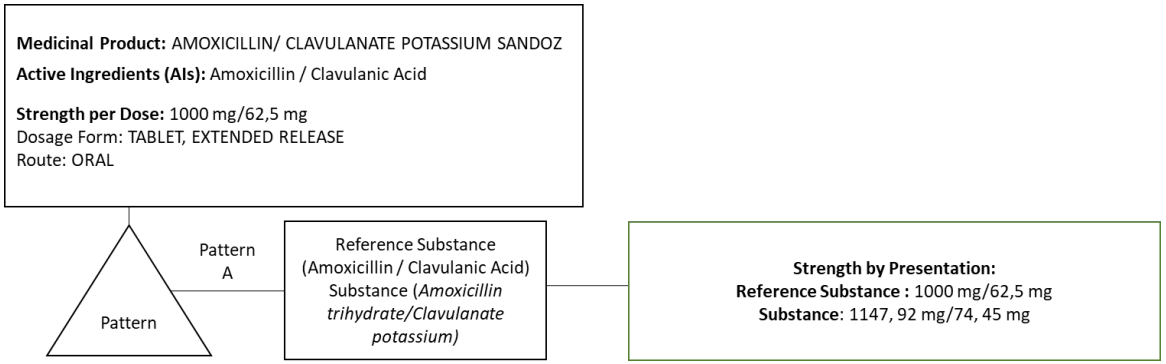
Pattern	**Type of product**	**Examples**	**Pharm Prod. unit of Present**	**Strength by Presentation**	**Strength by Concentration**
A	Solid, countable Solid dose forms in "container" Metered dose delivered by a metered actuation	Tablets, capsules, suppositories Powder or granules in sachet, Inhalers, Spray	Tablet, capsule Container (sachet, etc.) Actuation (inhaler, etc.)	Mandatory	Empty

B	Unit dose or continuous presentation (dosing is individual/not accurate and the total volume in the container is of less importance for dosing purposes)	Vials, Unit dose solutions, parenteral liquid, unit dose nebuliser solutions Bulk powders/granules, semi- solids "bulk" liquids (i.e. eye drops), spray that is not metered dose	N/A since it is the concentration that is relevant	Empty*	Mandatory
C	Products enclosed in a "presentation", where the dose has a delivery rate	Transdermal patches	Patch	Empty	Mandatory– as a delivery rate over time
D	Unit dose or continuous presentation, diluted to different final concentration depending on application such as Injection and/or infusions	Vials, parenteral liquid	Container (vial, etc.)	Mandatory	Mandatory

*Note that in order to simplify for PhPID generation the approach is to express the strength as Strength per Concentration only, ignoring expressing strength according to unit dose.

Please find below examples on patterns and their application by product:

Example 1: Pattern A - AMOXICILLIN/ CLAVULANATE POTASSIUM SANDOZ



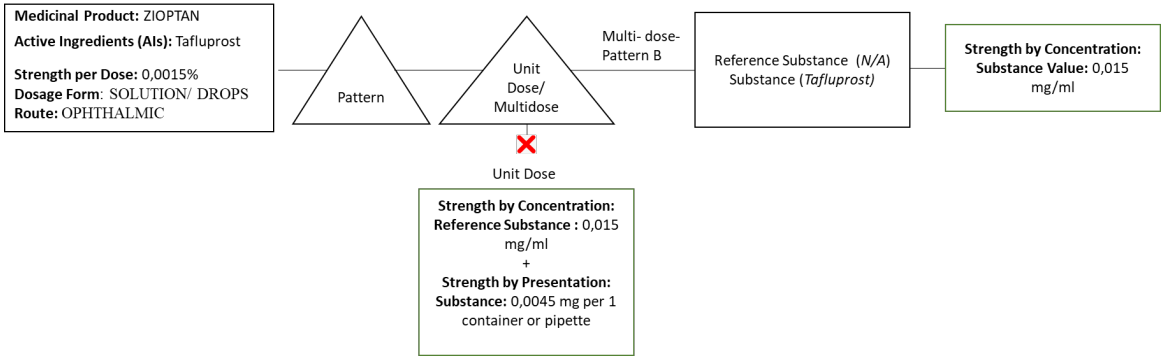
1. Based on Dose Form core characteristics (Release, Intended Site, Administration Method and Basic Administrable Dose Form) and other relevant characteristics (i.e. route) select pattern (see pattern framework above) – in this case the pattern to be selected is A.

2. In this case the reference substances of AIs (Amoxicillin / Clavulanic Acid) are expressed as Strength by Presentation. As outlined in ISO 11616, in order to unambiguously link the strength to the product - both strengths (reference substance (base) strength and substance (salt) strength) are deemed. It should be always verified in the SmPC if strength corresponds to substance or reference substance. In this case, the reference strength is derived from active moieties of an active substances(s) and corresponds to 1000 mg/62,5 mg. Since in the SmPC the active ingredients are given as substances (salts) the strength is expressed as (Amoxicillin trihydrate/Clavulanate potassium) as 1147,92 mg/74,45 mg, which is being calculated as follows:

   - amoxicillin 1000 mg/(amoxicillin base molecular weight 365,4 g/mol* amoxicillin trihydrate molecular weight 419,45 g/mol)=1147,92 mg;

   - Clavulanic acid 62,5 mg/(Clavulanic acid molecular weight 199,2g/mol* Clavulanate potassium molecular weight 237,25g/mol)=74,45 mg.

Example 2: Pattern B - ZIOPTAN



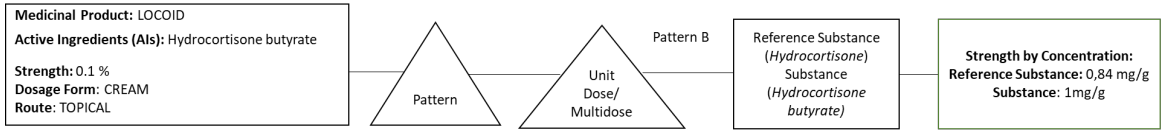
1. Based on Dose Form core characteristics (Release, Intended Site, Administration Method and Basic Administrable Dose Form) and other relevant characteristics (i.e. route) select pattern (see pattern framework above) – in this case the pattern to be selected is B

2. Double-check in the SmPC if strength corresponds to substance or reference substance: in this case the substance of AI is identical to reference substance (Tafluprost) and is expressed as Strength by Concentration: 0,015 mg/ml.

   *Note that the strength can be expressed as Strength per Presentation and as Strength per Concentration, depending on if it is a multi-dose package or unit dose package. In order to simplify for PhPID generation the recommendation is to express the strength as Strength per Concentration only, ignoring expressing strength according to unit dose.

3. The concentration is stated as 0.0015 % from FDA. Verify the strength in the SmPC. In the case above, both 0.0015 % and 0.015 mg/ml is stated in the reference. Priority to mg/ml according to EMA document https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/product-management-service-pms-implementation-international-organization-standardization-iso_en.pdf

Example 3: Pattern B - LOCOID:

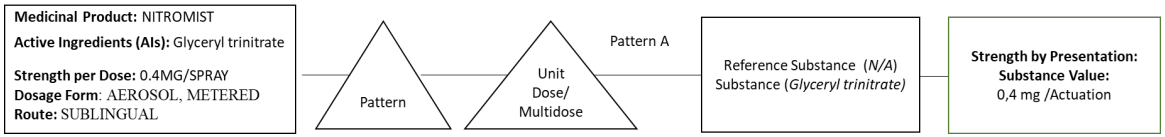


1. Based on based on Dose Form core characteristics (Release, Intended Site, Administration Method and Basic Administrable Dose Form) and other relevant characteristics (i.e. route) select pattern (see pattern framework above) – in the case above the pattern to be selected is B

2. Double-check in the SmPC if strength corresponds to substance or reference substance: in this case the substance of AI is Hydrocortisone butyrate 1 mg/g and the reference substance is calculated to be 0,84mg/g, and is expressed as Strength by Concentration: 1 mg/g. In regards to reference strength vs strength calculation is carried out as follows:

    - Given strength 1mg/g corresponds to the substance (salt) that is also indicated in the SmPC so the reference strength has to be calculated: hydrocortisone butyrate 1 mg/g (hydrocortisone butyrate salt molecular weight 432,55g/mol* hydrocortisone molecular weight 362,46g/mol)=0,84 mg;

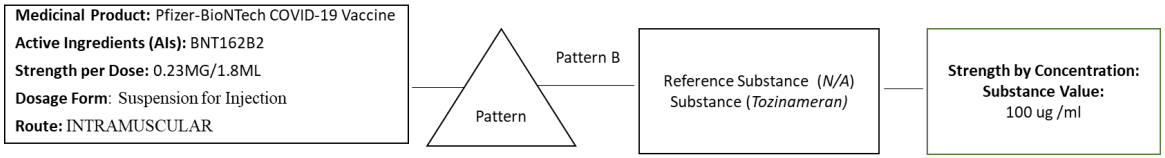
Example 4: Pattern A - NITROMIST:



1. Based on Dose Form core characteristics (Release, Intended Site, Administration Method and Basic Administrable Dose Form) and other relevant characteristics (i.e. route) select pattern (see pattern framework above) – in the case above the pattern to be selected is A

2. Double-check in the SmPC if strength corresponds to substance or reference substance: in this case the substance of AI is identical to reference substance (Glyceryl trinitrate) and is expressed as Strength by Presentation: 0,4 mg/Actuation.

Example 5: Pattern B



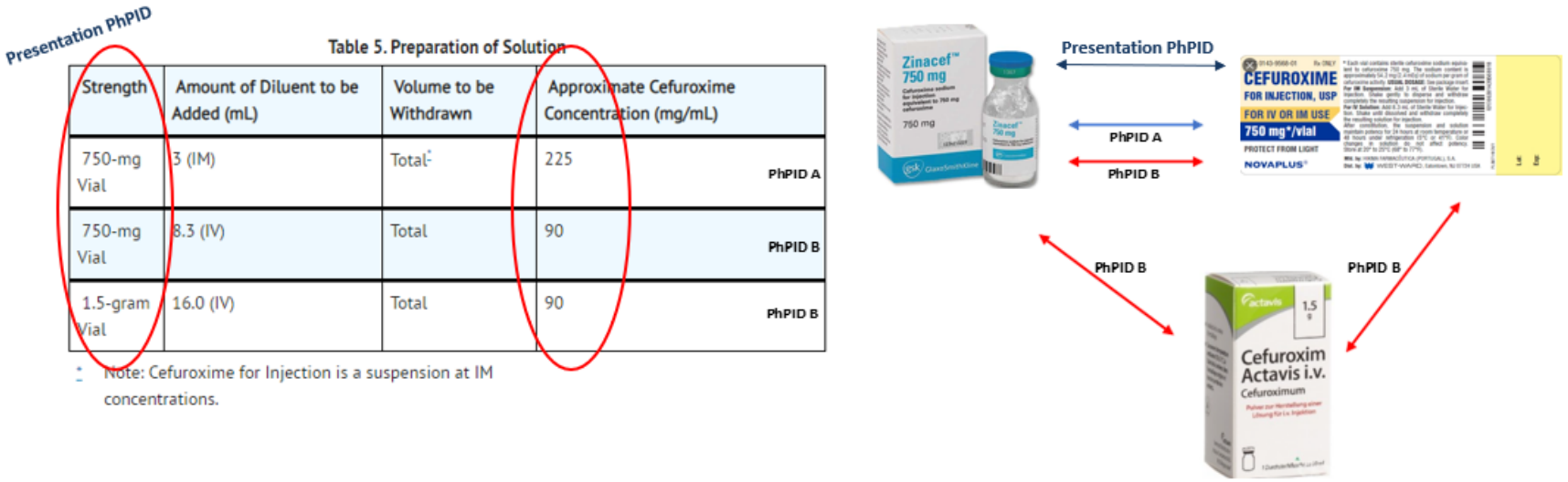
1. Based on based on Dose Form core characteristics (Release, Intended Site, Administration Method and Basic Administrable Dose Form) and other relevant characteristics (i.e. route) select pattern (see pattern framework above) – in the case above the pattern to be selected is B

2. Double-check in the SmPC if strength corresponds to substance or reference substance: in this case the substance of AI is identical to substance (Tozinameran) and is expressed as Strength by Concentration: 100 ug/ml.

Examples of application of different patterns and consequences on PHPID.

For liquid products, applying pattern B will in many cases result in aggregation PHPID for different qualitative and quantitative compositions while pattern D result in unique PHPID for most compositions.

Figure 2: Challenge with the products that have one presentation strength and two different concentration strength (i.e. Zinacef)



ZINACEF – Pattern D									
Preparation of Solution and Suspension				**Strength**		**EDQM dose form**	**Administration method**	**Pattern B PHPID by concentration only**	**Pattern D PHPID by presentation + concentration**
Strength	**Amount of Diluent Added (mL)**	**Volume to be Withdrawn**	**Approximate Concentration (mg/mL)**	**Presentation strength**	**Concentration strength**				
750-mg Vial	3.0 (IM)	Total	225	750 mg/vial	225 mg/ml	Solution for injection/ infusion	injection	YYZZ	XXYYZZ

750-mg Vial	8.3 (IV)	Total	90	750 mg/vial	90 mg/ml	Solution for injection/ infusion	infusion	UUVV	XXUUVV
1.5-gram Vial	16.0 (IV)	Total	90	1.5 g/vial	90 mg/ml	Solution for injection/ infusion	infusion	UUVV	WWUUVV

Zinacef is supplied as a dry powder in vials , in two different concentrations, where amount of diluent to be added, comes in three different volumes. Applying pattern B will result in aggregation into two PHPIDs, while pattern D result in unique PHPID for all above compositions.

SANDIMMUNE -Pattern B range expressed									
Concentrate for Infusion				**Strength**		**EDQM dose form**	**Administration method**	**Pattern B PHPID by concentration only**	**Pattern D PHPID by presentation + concentration**
Strength	**Amount of Diluent to be Added (mL)**	**Volume to be Withdrawn**	**Approximate Concentration (mg/mL)**	**Presentation strength**	**Concentration strength**				
50mg/mL	1+20 mL (IV/IM)	Total	2,5	250mg/ampul	2,5mg/mL	Solution for infusion	infusion	YYZZ	XXYYZZ
50mg/mL	1+100 mL (IV)	Total	0,5	250mg/ampul	0,5mg/mL	Solution for infusion	infusion	UUZZ	XXUUZZ

Sandimmune is supplied as a concentrate in ampuls, where amount of diluent to be added, can vary from 20 mL to 100 mL depending on clinical evaluation. Applying pattern B will result in aggregation into two PHPID, while pattern D result in unique PHPID for the above compositions.

LOVENOX – Pattern B and/or D									
Solution for Injection (prefilled syringes)				**Strength**		**EDQM dose form**	**Administration method**	**Pattern B PHPID by concentration only**	**Pattern D PHPID by presentation + concentration**
Strength	**Amount of Diluent to be Added (mL)**	**Volume to be Withdrawn**	**Approximate Concentration (mg/mL)**	**Presentation strength**	**Concentration strength**				
30mg/0,3mL	N/A	Total	100	30mg/prefilled syringe	100 mg/1mL	Solution for injection	injection	XXYY	AAXXYY
40mg/0,4mL	N/A	Total	100	40mg/prefilled syringe	100 mg/1mL	Solution for injection	injection	XXYY	BBXXYY
60mg/0,6mL	N/A	Total	100	60mg/prefilled syringe	100mg/1mL	Solution for injection	injection	XXYY	CCXXYY
120 mg/0.8 mL	N/A	Total	150	120 mg/syringe	150mg/1mL	Solution for injection	Injection	WWYY	DDWWYY
100mg/1ml	N/A	Total	100	100 mg/ampoule	100mg/1ml	Solution for injection	Injection	XXYY	EEXXYY

Lovenox is supplied as prefilled syringes and ampoules, in two different concentrations. Applying pattern B will result in aggregation into two PHPID, while pattern D result in unique PHPID for all above compositions.

SOLU-CORTEF- Pattern B range expressed									
Powder for Solution (vials)				**Strength**		**EDQM dose form**	**Administration method**	**Pattern B PHPID by concentration only**	**Pattern D PHPID by presentation + concentration**
Strength	**Amount of Diluent Added (mL)**	**Volume to be Withdrawn**	**Approximate Concentration (mg/mL)**	**Presentation strength**	**Concentration strength range**				

100mg/2mL	2mL +100 to 1000 mL	Total	0,1	100mg/vial	0,1mg/mL-0,99mg/mL	Solution for infusion	Infusion	YYZZ	AAYYZZ
250mg/2mL	2mL+250 to 1000 mL	Total	0,25	250mg/vial	0,1mg/mL-0,99mg/mL	Solution for infusion	infusion	YYZZ	BBYYZZ
500mg/4mL	4mL+500 to 1000mL	Total	0,5	500mg/vial	0,1mg/mL-0,99mg/mL	Solution for infusion	infusion	YYZZ	CCYYZZ
1000mg/8mL	8mL +1000mL	Total	0,99	1000mg/vial	0,1mg/mL-0,99mg/mL	Solution for infusion	infusion	YYZZ	DDYYZZ

Solu-cortef is supplied as single-dose vials in four different concentrations. Applying pattern B will result in aggregation into one PHPID, while pattern D result in unique PHPID for all above compositions.

CALDOLOR – Pattern D									
Solution for Injection (vials)				**Strength**		**EDQM dose form**	**Administration method**	**Pattern B PHPID by concentration only**	**Pattern D PHPID by presentation + concentration**
Strength	**Amount of Diluent to be Added (mL)**	**Volume to be Withdrawn**	**Approximate Concentration (mg/mL7)**	**Presentation strength**	**Concentration strength**				
100mg/mL	TBC	TBC	4mg/mL	800mg/vial	4mg/mL	Solution for Injection	Injection	YYZZ	AAYYZZ
4mg/mL	N/A	TBC	4mg/mL	800mg/bag	4mg/mL	Solution for Injection	Injection	YYZZ	BBYYZZ

The product Caldolor comes in two strength where the 100 mg/ml vial is diluted before use to the same concentration as the 4 mg/ml vial. Applying pattern D to these products would result in two unique PHPIDs while pattern B generate identical PHPID.

Generally, it could be discussed how to take dilution into account for PHPID and which strength expressions should be used. Some products are diluted sequentially and/or according to weight, resulting in many different final concentrations. The effort to find information about the dilutions in the SmPCs are substantial and a lot of manual calculations needed to express strength for the different dilutions. Including dilution in the strength expression has an impact on the PHPIDs and needs to be expressed consistently for harmonized PHPID generation.

Advantages with pattern B is aggregation of data to get an enlarged dataset for evaluation while pattern D results in unique PHPIDs which could ensure easier identification of for example medication errors.

The procedure for presentation to the chosen MD5 HASH function

Order of input data

The order for input into the MD5 hash function was: substance/ strength/ dose form for this pilot to keep the substance information with its relating strength (Note: strength values have been rounded to two decimal places).

Example of numerical representations

Table 7. Numerical representation of 3 Covid -19 vaccine using the UMC simplified substance ID and EDQM characteristics codes. The reference substance and reference strength is identical to substance and strength in this case and have been excluded from the table.

Product	Substance ID	Strength by presentation	Strength by concentration	Basic Dose Form code	Administration Method Form Code	Intended Site Code	Release Characteristics Code
Covid-19 vaccine Pfizer (Comirnaty)	36667	NA	100 µg/ml	0085	0011	0022	0047
Covid-19 vaccine Moderna	35854	NA	200 µg/ml	0085	0011	0022	0047
Covid-19 vaccine AstraZeneca	35853	NA	5×10^8 infectious units	0085	0011	0022	0047

Outcome analysis

The purpose of this section is to analyze the PhPIDs generated and understand if a selected set of dose form characteristics from EDQM and other potential characteristics can describe dose forms and be utilized to solve the issues with mapping between different dose form terminologies.

In order to perform a thorough outcome analysis, the following methods were applied:

Dose Form Mapping exercise and Findings

In order to investigate how using Release, Intended Site, Administration Method and Basic Administrable Dose Form would affect harmonization of Dose Forms - EDQM Administrable Dose Form were mapped to Dose Form core characteristics.

Please find below some examples and related findings on cases where PhPIDs will be the same using core characteristics even if there are differences in EDQM Adminstrable Dose Forms.

TABLETS

PhPID	EDQM Administrable Dose Form	Basic Dose Form	Administration Method	Intended Site	Release Characteristics
Same PhPID	Sublingual tablet	Tablet	Administration	Oromucosal	Conventional
	Buccal tablet	Tablet	Administration	Oromucosal	Conventional
Unique PhPID	Muco-adhesive buccal tablet	Tablet	Application	Oromucosal	Prolonged
Unique PhPID	Chewable tablet	Tablet	Chewing	Oral	Conventional
Unique PhPID	Chewable/dispersible tablet	Tablet	Chewing/ Swallowing	Oral	Conventional
Unique PhPID	Implantation tablet	Tablet	Implantation	Parenteral	Prolonged
Unique PhPID	Inhalation powder, tablet	Tablet	Inhalation	Pulmonary	Conventional
Same PhPID	Vaginal tablet	Tablet	Insertion	Vaginal	Conventional
	Effervescent vaginal tablet	Tablet	Insertion	Vaginal	Conventional
Unique PhPID	Orodispersible tablet	Tablet	Orodispersion	Oral	Conventional
Same PhPID	Tablet	Tablet	Swallowing	Oral	Conventional
	Coated tablet	Tablet	Swallowing	Oral	Conventional
	Film-coated tablet	Tablet	Swallowing	Oral	Conventional
	Tablet with sensor	Tablet	Swallowing	Oral	Conventional
Unique PhPID	Gastro-resistant tablet	Tablet	Swallowing	Oral	Delayed
Unique PhPID	Prolonged-release tablet	Tablet	Swallowing	Oral	Prolonged
Unique PhPID	Modified-release tablet	Tablet	Swallowing	Oral	Modified

Finding: SmPC does not always provide the details on administrable dose form.

EDQM Administrable Dose Form such as tablet, coated tablet, film-coated tablet, and tablet with sensor will have same PhPID cased on chosen characteristics.

EDQM Administrable Dose Form mapping for sublingual tablet and buccal tablet will also result in same PhPID.

From pharmacovigilance perspective generating same PhPIDs for the above EDQM terms would be beneficial. From prescription perspective this approach would represent a potential risk for the patient receiving inappropriate pharmaceutical form.

CAPSULES

PhPID	EDQM Administrable Dose Form	Basic Dose Form	Administration Method	Intended Site	Release Characteristics
Unique PhPID	Chewable capsule, soft	Capsule	Chewing	Oral	Conventional
Unique PhPID	Oromucosal capsule	Capsule	Chewing	Oromucosal	Conventional
Unique PhPID	Inhalation vapour, capsule	Capsule	Inhalation	Pulmonary	Conventional
Same PhPID	Vaginal capsule, hard	Capsule	Insertion	Vaginal	Conventional
	Vaginal capsule, soft	Capsule	Insertion	Vaginal	Conventional
Unique PhPID	Rectal capsule	Capsule	Insertion	Rectal	Conventional
Same PhPID	Capsule, hard	Capsule	Swallowing	Oral	Conventional
	Capsule, soft	Capsule	Swallowing	Oral	Conventional
Same PhPID	Gastro-resistant capsule, hard	Capsule	Swallowing	Oral	Delayed
	Gastro-resistant capsule, soft	Capsule	Swallowing	Oral	Delayed
Same PhPID	Prolonged-release capsule, hard	Capsule	Swallowing	Oral	Prolonged
	Prolonged-release capsule, soft	Capsule	Swallowing	Oral	Prolonged
Same PhPID	Modified-release capsule, hard	Capsule	Swallowing	Oral	Modified
	Modified-release capsule, soft	Capsule	Swallowing	Oral	Modified

Findings: SmPC does not always provide the details on if capsule is hard or soft.

From pharmacovigilance perspective generating same PhPIDs for the above EDQM Administrable Dose Form would be beneficial grouping EDQM Administrable Dose Form should group “hard” and “soft” capsules together. From prescription perspective this approach would represent a potential risk for the patient receiving inappropriate pharmaceutical form.

SOLUTION – INFUSION/INJECTION

PhPID	Administrable Dose Form	Basic Dose Form	Administration Method	Intended Site	Release Characteristics
Unique PhPID	Dispersion for injection/infusion	Dispersion	Infusion/injection	Parenteral	Conventional
Unique PhPID	Dispersion for infusion	Dispersion	Infusion	Parenteral	Conventional
Unique PhPID	Emulsion for infusion	Emulsion	Infusion	Parenteral	Conventional
Unique PhPID	Emulsion for injection/infusion	Emulsion	Infusion/injection	Parenteral	Conventional
Unique PhPID	Solution for infusion	Solution	Infusion	Parenteral	Conventional
Unique PhPID	Solution for injection/infusion	Solution	Infusion/injection	Parenteral	Conventional
Unique PhPID	Solution for injection	Solution	Injection	Parenteral	Conventional
Unique PhPID	Suspension for injection	Suspension	Injection	Parenteral	Conventional
Unique PhPID	Emulsion for injection	Emulsion	Injection	Parenteral	Conventional
Unique PhPID	Gel for injection	Gel	Injection	Parenteral	Conventional
Unique PhPID	Prolonged-release suspension for injection	Suspension	Injection	Parenteral	Prolonged
Unique PhPID	Intraperitoneal solution	Solution	Injection	Intraperitoneal	Conventional
Unique PhPID	Prolonged-release solution for injection	Solution	Injection	Parenteral	Prolonged
Unique PhPID	Prolonged-release dispersion for injection	Dispersion	Injection	Parenteral	Prolonged
Unique PhPID	Dispersion for injection	Dispersion	Injection	Parenteral	Conventional

All dose forms result in unique PhPIDs.

SOLUTION – ORAL SUSPENSION/ORAL SOLUTION

PhPID	EDQM Administrable Dose Form	Basic Dose Form	Administration Method	Intended Site	Release Characteristics
Unique PhPID	Oral suspension	Suspension	Swallowing	Oral	Conventional
Unique PhPID	Gastro-resistant oral suspension	Suspension	Swallowing	Oral	Delayed
Unique PhPID	Prolonged-release oral suspension	Suspension	Swallowing	Oral	Prolonged
Unique PhPID	Modified-release oral suspension	Suspension	Swallowing	Oral	Modified
Unique PhPID	Oral drops, solution	Solution	Instillation/ Swallowing	Oral	Conventional
Unique PhPID	Oral solution	Solution	Swallowing	Oral	Conventional
Unique PhPID	Oral/rectal solution	Solution	Administration/ Swallowing	Oral/Rectal	Conventional
Unique PhPID	Oral solution/concentrate for nebuliser solution	Solution	Inhalation/ Swallowing	Oral/Pulmonary	Conventional

Finding: In some cases, delivery devices represent the only way to understand if a product is a solution or a suspension according to EDQM. Detailed information on delivery device (syringe for suspension/spoon for solution) is not always described explicitly in SmPC. Clarification in EDQM could potentially improve understanding and assignment of these dose forms.

CREAMS vs OINTMENTS

PhPID	EDQM Pharmaceutical Administrable Dose Form	Basic Dose Form	Administration Method	Intended Site	Release Characteristics
Unique PhPID	Ear ointment	Ointment	Application	Auricular	Conventional
Unique PhPID	Ear/eye ointment	Ointment	Application	Auricular/Ocular	Conventional
Unique PhPID	Oromucosal cream	Cream	Application	Oromucosal	Conventional
Unique PhPID	Cream	Cream	Application	Cutaneous/Transdermal	Conventional
Unique PhPID	Eye cream	Cream	Application	Ocular	Conventional
Unique PhPID	Ear cream	Cream	Application	Auricular	Conventional

Unique PhPID	Nasal cream	Cream	Application	Nasal	Conventional
Unique PhPID	Vaginal cream	Cream	Application	Vaginal	Conventional
Unique PhPID	Rectal cream	Cream	Application	Rectal	Conventional
Same PhPID	Ointment	Ointment	Application	Cutaneous/Transdermal	Conventional
	Transdermal ointment	Ointment	Application	Cutaneous/Transdermal	Conventional
	Cutaneous/nasal ointment	Ointment	Application	Cutaneous/Transdermal	Conventional
Unique PhPID	Cutaneous spray, ointment	Ointment	Spraying	Cutaneous/Transdermal	Conventional
Unique PhPID	Urethral ointment	Ointment	Administration	Intravesical/Urethral	Conventional
Unique PhPID	Nasal ointment	Ointment	Application	Nasal	Conventional
Unique PhPID	Eye ointment	Ointment	Application	Ocular	Conventional
Unique PhPID	Oromucosal ointment	Ointment	Application	Oromucosal	Conventional
Unique PhPID	Inhalation vapour, ointment	Ointment	Inhalation	Pulmonary	Conventional
Unique PhPID	Rectal ointment	Ointment	Application	Rectal	Conventional
Unique PhPID	Vaginal ointment	Ointment	Application	Vaginal	Conventional

Finding: SmPC does not always provide the details on administrable dose form.

EDQM Administrable Dose Form mapping for Ointment, Transdermal ointment and Cutaneous/nasal ointment resulting in same PhPID.

From pharmacovigilance perspective generating same PhPIDs for the above EDQM terms would be beneficial. From prescription perspective this approach would represent a potential risk for the patient receiving inappropriate pharmaceutical form

PATCH

PhPID	EDQM Pharmaceutical Administrable Dose Form	Basic Dose Form	Administration Method	Intended Site	Release Characteristics
Unique PhPID	Cutaneous patch	Patch	Application	Cutaneous/Transdermal	Conventional
Unique PhPID	Transdermal patch	Patch	Application	Cutaneous/Transdermal	Prolonged
Unique PhPID	Oromucosal patch	Patch	Application	Oromucosal	Prolonged

Finding: Though EDQM definitions are quite straightforward on both Cutaneous and Transdermal - what is rationale behind to have dual intended site?

The curent dwefinitions in EDQM:

Cutaneous patch - Flexible single-dose preparation intended to be applied to the unbroken skin to obtain a local effect by penetration of the active substance(s) into the skin.

Transdermal patch - Flexible single-dose preparation intended to be applied to the unbroken skin to obtain a systemic delivery over an extended period of time. Transdermal patches consist of a backing sheet supporting a reservoir or a matrix containing the active substance(s) and on the top a pressure-sensitive adhesive, which assures the adhesion of the preparation to the skin. The backing sheet is impermeable to the active substance(s) and normally impermeable to water. In reservoir systems the active substance may be dissolved or dispersed in a semi-solid basis or in a solid polymer matrix, which is separated from the skin by a rate-controlling membrane. The pressure-sensitive adhesive may, in this case, be applied to some or all parts of the membrane, or only around the border of the membrane and the backing sheet. Matrix systems contain the active substance in a solid or semi-solid matrix, the properties of which control the diffusion pattern to the skin. The matrix system may also be a solution or dispersion of the active substance in the pressure-sensitive adhesive. The releasing surface of the patch is covered by a protective liner to be removed before applying the patch to the skin.

OCULAR FORMULATIONS

PhPID	EDQM Pharmaceutical Administrable Dose Form	Basic Dose Form	Administration Method	Intended Site	Release Characteristics
Unique PhPID	Eye cream	Cream	Application	Ocular	Conventional
Unique PhPID	Eye gel	Gel	Application	Ocular	Conventional
Unique PhPID	Eye ointment	Ointment	Application	Ocular	Conventional
Unique PhPID	Eye drops, solution	Solution	Instillation/ Swallowing	Ocular	Conventional
Unique PhPID	Eye drops, emulsion	Emulsion	Instillation/ Swallowing	Ocular	Conventional
Unique PhPID	Eye drops, suspension	Suspension	Instillation/ Swallowing	Ocular	Conventional
Unique PhPID	Eye drops, prolonged-release	Drops (unspecified)	Instillation	Ocular	Prolonged
Unique PhPID	Eye lotion	Solution	Bathing	Ocular	Conventional

Finding: SmPC does not always provide the details on administrable dose form if it is a solution/emulsion/suspension, which can cause PhPID inconsistency.

COVID-19 VACCINE DOSE FORMS

Finding: COVID-19 vaccines dose form assignment varies within different authorities. EDQM definition of the different dose forms used could be improved to simplify a harmonized assignment. Furthermore, when dose forms are expressed differently within different jurisdictions, a clear base for decision of which dose form should be used to create a global PHPID.

Please find example of dose forms assigned for the Covid-19 vaccine Comirnaty within different authorities:

Authority of Approval	Administrable Dose Form	EDQM Definition
EMA	dispersion for injection	Liquid sterile preparation consisting of two or more phases of which at least one is dispersed in the liquid phase, intended for administration by injection. To be used only when emulsion for injection is not appropriate. Solid suspension preparations are excluded.
FDA	suspension for injection	Liquid sterile single-dose or multidose preparation consisting of a suspension intended for administration by injection
UK	solution for injection	Liquid sterile single-dose or multidose preparation consisting of a solution intended for administration by injection.

Conclusion & Recommendations

This section outlines the 'Recommendations/Improvement Suggestions' coming from the pilot results as well as the high-level 'Action Plan' activities proposed for process/system owners and/or key stakeholders.

Pilot results confirmed that centralized core EDQM dose form characteristics (and their codes), Release Characteristics, Intended Site, Administration Method and Basic Administrable Dose Form, can be used as input in the generation of the global PhPID and solve the issues with mapping between different dose form terminologies.

Using centralized EDQM core characteristics would harmonize the levels of granularity between regions and significantly increase the quality of one-to-one mapping between a regional terminology and a centrally controlled vocabulary. This approach would ensure a consistent PhPID construction and thereby allow gathering information into one global data source.

Having a global data repository would facilitate a much faster and more efficient detection of drug safety signals, and substantially increase the probability of detecting rare adverse drug reactions (ADRs), proving basis for holistic pharmacovigilance support and enabling exchange of ICSR information between NCs/regulators, industry and other stakeholders globally.

However, the pilot results have demonstrated also that there are challenges to overcome and there is no "perfect" solution.

The findings demonstrated PhPID generation issues described below by section.

Dose Forms Issues/Challenges

- Cases when dose forms are expressed differently within different jurisdictions
- Cases where certain dose form characteristic have multiple values
- Cases when medicinal product dose form description is twofold in SmPC

Recommendation/Improvement Suggestions

Regulatory agencies

Regulatory agencies can maintain their regional terminologies, while using centralized dose form characteristics for global PhPID.

Regulatory agencies can use the approach prospectively and retrospectively to assign the characteristics to new and to currently marketed medicinal products.

Global organisations

Global organizations can assign dose form characteristics to regional terminologies which is beneficial for global implementation of PHPID.

Best practices of using centralized core dose form characteristics and other potential characteristics for describing dose forms needs to be developed.

Some clarification in EDQM could potentially improve understanding and assignment of dose form characteristics.

Substance Issues/Challenges

The substance descriptions in the SmPC/label were sometimes inadequate and the detail of information varied, particular with regards to hydrates. There is a discrepancy between the naming of hydrates in different pharmacopieas and/or other naming bodies around the world. Information about hydrate variation of a substance is most often not listed as part of the active ingredient on packages).

Recommendation/Improvement Suggestions

Global organisations

Agree on global PhPID submission criteria/requirements for substances and develop a best practice to be used when applying for PhPID

Strength Issues/Challenges

Strength expression variations

- Variation in use of units for strength expression for similar products
- When strength is expressed differently within different jurisdictions

Hydrates- Label information variations affecting strength expression

- The generation of the PHPID is affected by hydrate since the strength needs to be calculated based the molecular weight which differs between hydrates and anhydrous chemicals.

Strength interval

- The use cases for expression of strength interval needs to be clarified.

Products requiring dilutions

- The strength expression for medicinal product formulated as a concentrate and which shall be diluted with an unknown amount of liquid needs to be clarified.
- Strength expression for products where different amounts of diluent is to be added, resulting in one presentation strength and two or more different concentration strengths needs to be discussed

Recommendation/Improvement Suggestions

Global organisations

Agree on a consistent way to express the strength for different types of products

Agree on harmonised strength expression for similar products and issue a guidance for strength information submission

Implement Strength Patterns Framework

Agree on and develop best practices of using information currently available on strength to ensure consistent PhPID assignments

Reference list

https://www.edqm.eu/sites/default/files/standard_terms_introduction_and_guidance_for_use.pdf

https://standardterms.edqm.eu/

Document version	D8.1_ANNEX 4_ANALYSIS_EDQM_TERMINOLOGY
Authors	J.Roumier R. Vander Stichele D. van Nimwegen
Changes	15/09/2021 - V0.6 Discussion with R. Vander Stichele – more details and examples 30/09/2021 - V1.0 final version

D8.1_Annex 5 Analysis of EDQM dose form terminology and proposal for dose form ontology

Description of the proposed changes to classic EDQM in the light of a transition to a global terminology

Contents

Introduction

The EDQM [European Directorate for the Quality of Medicines & Healthcare] Standard Terms database contains terms and definitions to describe pharmaceutical dose forms, routes and methods of administration, containers, closures, administration devices and units of presentation. This is an important and sound resource. Documented here are an analysis of the database and a series of small modifications proposed to allow the creation of a global terminology.

Analysis of the EDQM Standard Terms database

We work in this document with the version of EDQM Standard Terms database [STD] of 2021-03-17 09:52:21. We do not provide a complete descripiton of STD but focus on the data categories that are relevant here.

With this version of STD the number of

Pharmaceutical Dose Forms [PDF] is 563. For the purpose of this analysis, we discarded a number of PDFs, only keeping 428 of them :

- domain
  - 79 with value « Veterinary only », only keeping the « Human and Veterinary »
- status
  - 23 with status « Rejected »
  - 37 with status « Deprecated »
  - 2 with status « Pending »
  - only keeping the « Current »

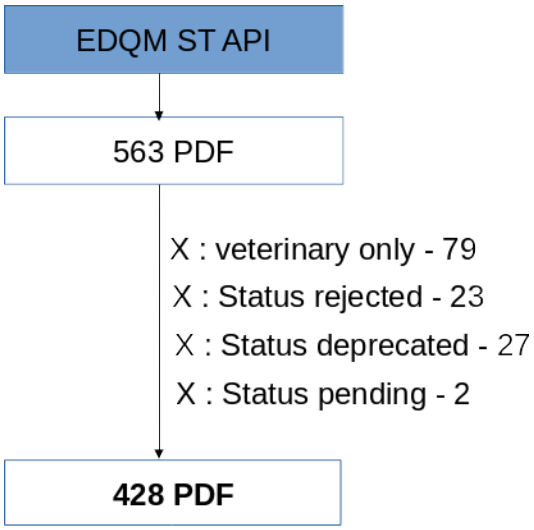


Figure 1: Selecting relevant PDFs

In our version of the STD, all PDFs are also tagged with ST [Standard Term] and a subset that are an administrable form with AdmDF : 295. Hence the number of PDFs that require at least a transformation : 131.

Other relevant data categories are :

- PDF code : 10101000
- english label : Oral drops, solution
- definition : Liquid, usually multidose preparation consisting of a solution intended for oral use. The preparation is administered in small volumes by means of a suitable measuring device such as a dropper, pipette or oral syringe capable of accurate dosing of the solution. The measured dose may be diluted in water or another suitable liquid before swallowing.
- Basic Dose Form [BDF] code and english term : 0083, Solution
- State of Matter [SOM] code and english term : 0099, Liquid
- [multiple] Transformation [TRA] code and english term : 0042, No transformation
- [multiple] Release Characteristics [RCA] code and english term : 0047, Conventional
- [multiple] Intended Site [ISI] code and english term : 0031, oral
- [multilple] Administration Method [AME] code and english term : 0013, Instillation

Initial critical analysis

After the surface analysis of the STD, a critical analysis lead to several observation for the improvement of the resource :

- The last four characteristics (TRA, RCA, ISI, AME) are also found in the English written definition of the PDFs and sometimes the characteristics and definitions are not perfectly aligned.
- The tags system of STD contains a tag « ST » that is always present hence does not convey new information. The AdmDF allows to easily identify PDF not directly administrable but not the potential transformation that lead to it. Also, the status of a PDF tagged with AdmDF is ambiguous when associated with a value different from « No transformation » for TRA. Sometimes this is clarified in the definition.
- Having mutltiple values for the four characteristics is visually helping but makes it difficult to programmatically use them, and leads to the need to concanate the values into new « combined » values and the creation of new codes.
- The ISI value Cutaneous/Transdermal is not split into two different ISI values. Splitting it would be clinically beneficial and the PDF definition most often indicate it is possible.
- The differenciation between « Systemic » and « Non-Systemic » effect would be clinically beneficial and PDF definitions most often indicate it is possible.

This series of small modifications would make STD more adequate for a global terminology of dose forms.

Micro-surgery

To make the benefits of the proposed modifications more tangible, they have been implemented in a new version, derived from the STD :

- The four characteristics (TRA, RCA, ISI, AME) can be multiple, and require the creation of additional « combined » characteristics that consist in the combination of the multiple values, e.g. Oral drops, solution (10101000) has 2 AME : Instillation(0013) and Swallowing (0019), leading to AME combined [AMEC] : instillation/swallowing(13/19). To avoid the « / » in the code, new codes were created marked with 999 to clearly differentiate them from the current EDQM codes : 999070000 for AMEC « instillation/swallowing. RCA does not have multiple values for the domain « Human » , so the RCAC was not developped. ISI and TRA lead respectively to ISIC and TRAC.

- Creation of a new characteristic « Administrable Dose Form » [ADF] to make explicit the result of the transformation of a PDF. A new associated code is created for the ADF that do not correspond exactly to an existing PDF.
  - For example, PDF « inhalation vapour, capsule (11113000) » is transformed to ADF « inhalation vapour », non existing as a PDF and for which a new code was created : »999010000 »
  - ADF often have BDF and SOM that do not correspond to the transformed PDF, requiring the creation of new characteristics : ADF_BDF and ADF_SOM. ADF_BDF and ADF_SOM have the same value sets BDF and SOM have.
  - For « Bath additive (10501000) » no clear ADF could be found.

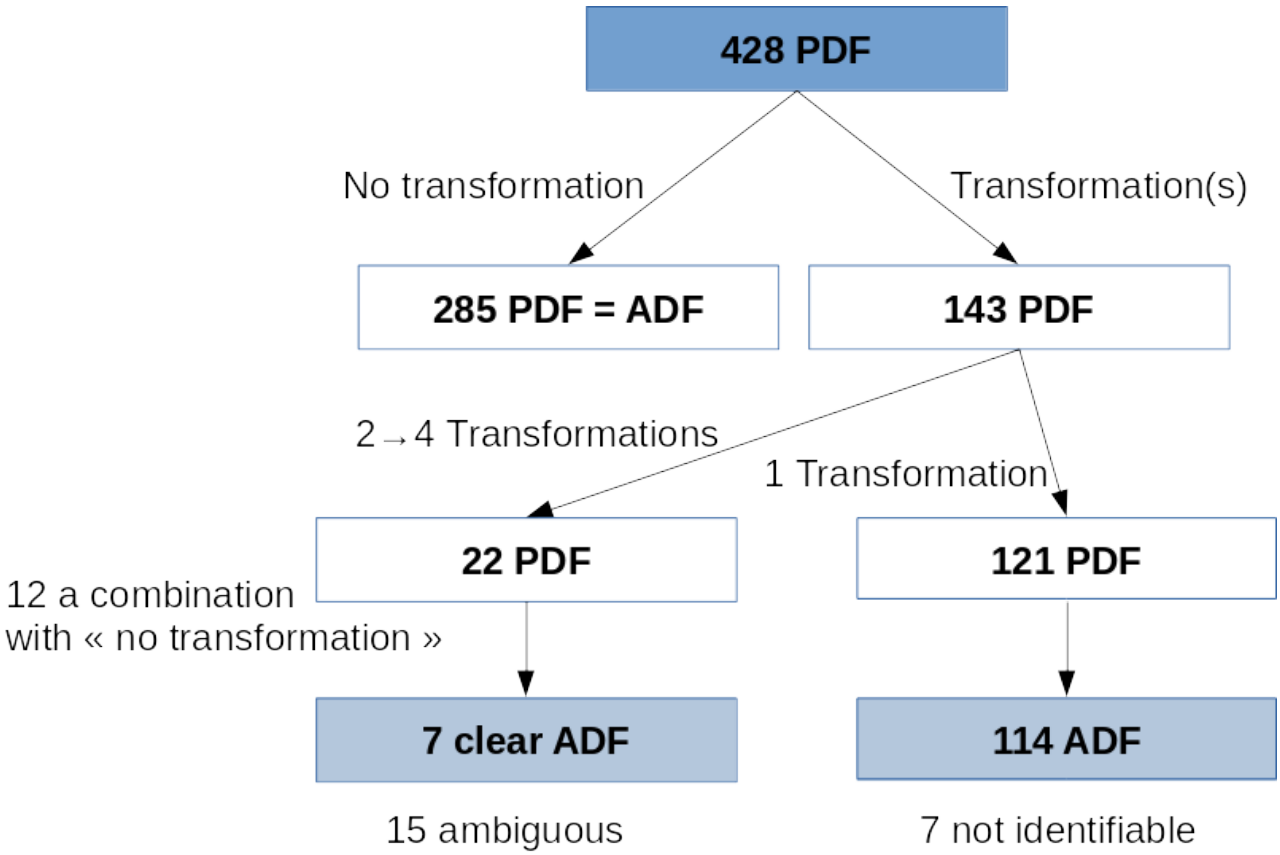


Figure 2: Finding ADFs from PDFs

- Splitting Cutaneous/Transdermal was possible for most. Examples :
  - « bath additive (10501000) » is now « cutaneous » instead of « cutaneous/transdermal »
  - « transdermal patch (10519000) » is now « transdermal » instead of « cutaneous/transdermal »
  - « medical larvae (13124000) » remains « cutaneous/transdermal »
  - All cutaneous PDF have a local effect, All transdermal have a systemic effect, and all « cutaneous/transdermal » PDFs have a local effect, except « cream (10502000) » which can have both local and systemic effect.
- Introducing the additional characteristic « Systemic/Non-systemic » was most of the time straight-forward based on the ISI, and relied on
  - The split of Cutaneous/Transdermal
  - Categorizing sublingual PDFs as Systemic
    - Currently the value-set of ISI for all sublingual PDFs is « oromucosal ». Having a new ISI value « sublingual » would reflect the fact that they are the only « oromucosal » with systemic effect where all others have a local effect. This has been implemented in the 3-level ontology described bellow.
  - Note : for some Nasal and Rectal dose forms, refinements are required to identify potential systemic PDF : Nasal spray, Rectal suppository.

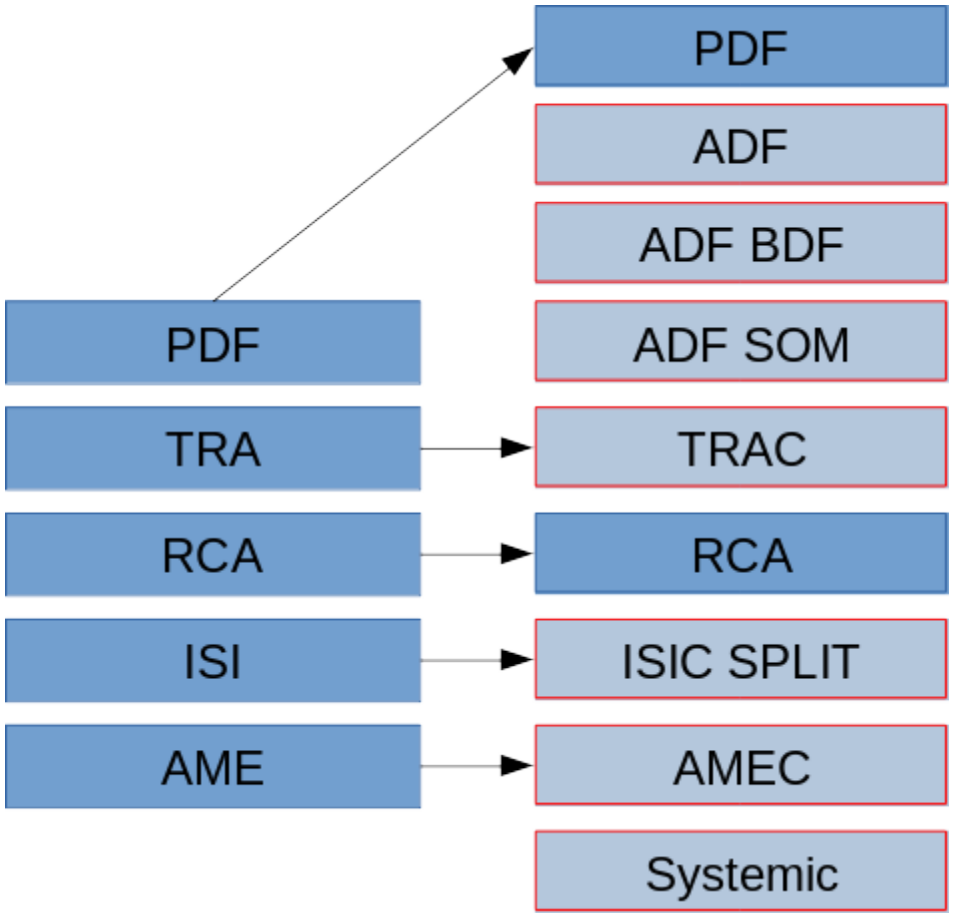


Figure 3: Modifications and additions to the orginal set of characteristics

The resulting characteristics set is close to be definitional, except for a couple of PDF

	ADF	BDF	SOM	ADF_BDF	ADF_SOM	TRAC	RCA	ISIC_SPLIT	AMEC	SYS	Count UC	Doubles>1	Sum Doubles
1	x	x	x	x	x	x	x	x	x	x	420	4	8
2	x	x	x	x	x	x	x	x	x		402	12	26
3	x			x	x	x	x	x	x	x	381	21	47
4	x			x	x	x	x	x	x		380	21	48
5	x					x	x	x	x	x	378	22	50
6	x					x	x	x	x		377	22	51
7		x	x	x	x	x	x	x	x	x	350	33	78
8		x	x	x	x	x	x	x	x		340	37	88
9				x	x	x	x	x	x	x	306	51	122
10				x	x	x	x	x	x		293	56	135
11						x	x	x	x	x	128	78	300
12						x	x	x	x		113	79	315
UMC/FDA+sys				x			x	x	x	x	197	77	231
UMC/FDA				x			x	x	x		179	82	249

Figure 4: Evaluation of the potential of EDQM characteristics to be definitional

These exceptions must be worked on (based on the defintions) but could also disappear by obsoleting the original PDFs in EDQM, when the importance of the difference is debatable. For example is the difference between « Capsule, hard » and « Capsule, soft » relevant ? A first analysis did not find necessary difference in the definitions. No requirement to change the definitions is made here, it's the responsibility of EDQM and these characteristics are used for important acts. The appearance of new non-definitional characteristics sets for PDFs could be avoided by screening them before adding them to STD, but this is also not a requirement.

Creating a dose form ontology to add a level of granularity

The global Dose Form terminology must be tailored to link to different existing Dose Form terminologies, some more high level, some more granular.

The main steps consisted in a spreadsheet to :

1. Making groups of identical 4 characteristics (TRAC, RCA, AMEC, ISIC_SPLIT)
2. Classifying the groups alphabetically by ISIC_SPLIT
3. Putting the PDFs with « no transformation » first

Then to concatenate or split the groups based on 2 main criterion : clinical relevance and impact on the business rules for the determination of the strength. These business rules will be determined by the WHO and the FDA, based on specifict requirements provided by the project.

Examples : the Auricular group consists in all the PDFs that differ for some characteristics but not in a way that is clinically relevant and can hence be concatenated. One group had identical characteristics but was splitted nonetheless : Oral drops and Oral liquids.

The dose form groups created were then named (e.g. Cutaneous dose form).

AURICULAR	
	Auricular local dose form
	Auricular/nasal dose form
	Auricular/nasal/ocular dose form
	Auricular/ocular dose form
CUTANEOUS	
	Cutaneous dose form
	Cutaneous/transdermal dose form
	Cutaneous/transdermal/nasal dose form
	Cuteneous/oromucosal dose form
	Cutaneous/transdermal/parenteral dose form
DENTAL	
	Dental dose form
ENDOCERVICAL	
	Endocervical dose form
EXTRACORPERAL	
	Extracorporal dose form
EXTRACORPORAL/PARENTERAL	
	Dose form for dialysis
GASTRIC	
	Gastric dose form
GASTROENTERAL	
	Gastroenteral dose form

Figure 5: Naming the dose form groups

In the resulting ontology the dose form groups are linked to the Intended Sites (e.g. Cutaneous) through the property hasDoseForm (with the corresponding property hasISI leading back to the Intended Site). The granular PDFs are all subclasses of the grouped Dose Forms.

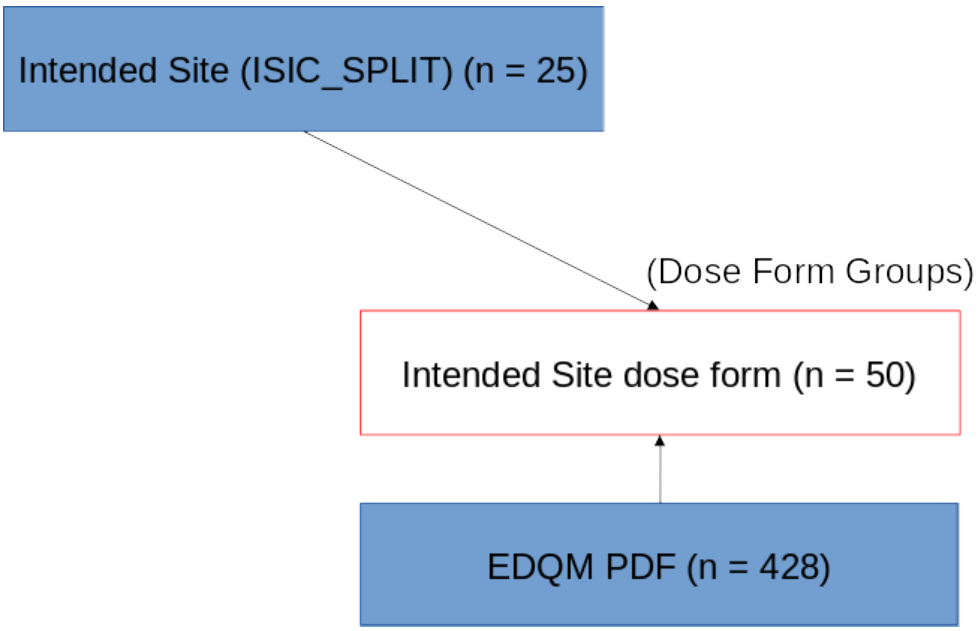


Figure 6: Structure of the simple dose form ontology

The ontology is available on WebProtégé, and is created automatically in OWL/XML from the source spreadsheet, based on a custom program in Java, OpenJena and OpenCSV.

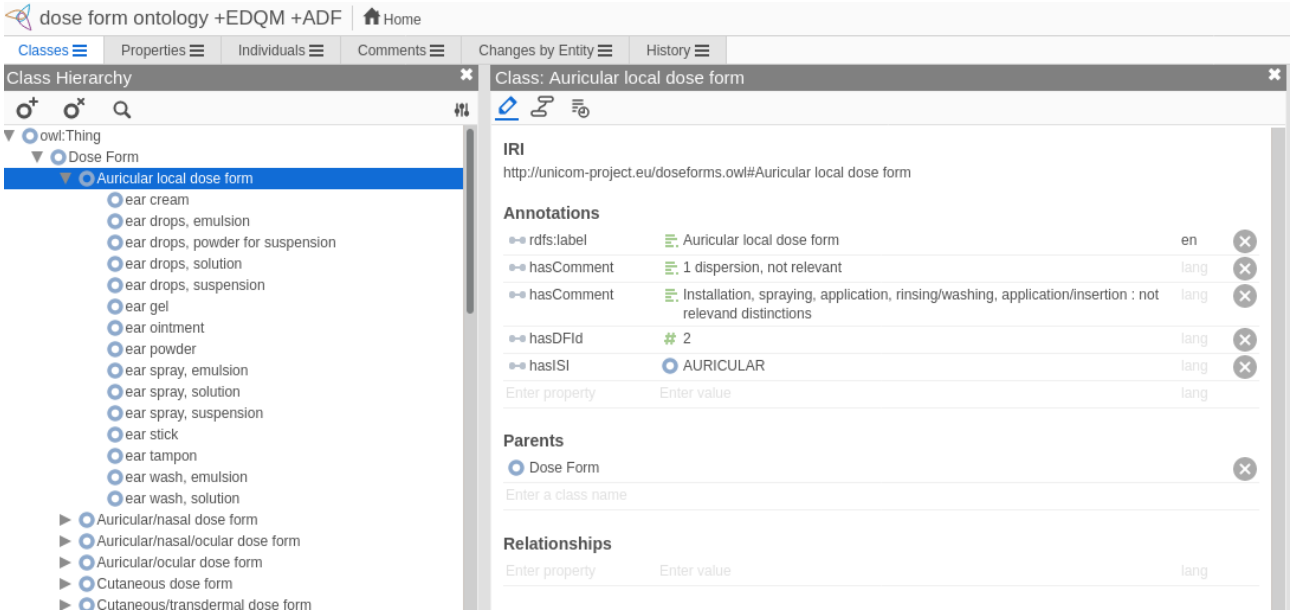


Figure 7: Visualising a PDF in the ontology (in WebProtégé)

Remarks:

- Some Oral and Rectal combinations still have to be solved. The combination should be kept as much as possible on the higher level.
- The « to be transformed dose form » ontological class was created as a temporary class for the study only. It contains two interesting properties :
  - hasADF links to the ADF obtained after transformation
  - a label containing in a text form the course of the transformation and the resulting ADF.
- Four isolated ADF arrive just below owl:thing, coming from faulty ADF not corresponding to existing PDF

In addition, this simple ontology has been used by Natalie J Karapetian to link the RxNorm to the dose form groups, allowing for further analysis of its usefulness.

Conclusion:

The EDQM STD is the best candidate for a global terminology of dose forms. Analysing the structure and content of the standard terms by the means of spreadsheets and ontologies allowed to discover and propose precise, surgical modifications to the resource to improve its capacity to be a central resource for the characterisation of Pharmaceutical dose forms. An additional step towards this was made with the creation of a simple and small ontology. By adding a middle layer of granularity – dose form groups – to the EDQM STD which is more granular than RxNorm for example, the alignement of dose forms between EDQM and other descriptions of dose forms used widely like SNOMED-CT can be improved. The work reported here will be continued in collaboration with UNICOM partners.

Document version	D8.1_ANNEX 5_Comparison EDQM/RxNORM
Authors	Natalie Karapetian Yuri Quintana Robert Vander Stichele
Changes	Contribution of Harvard Medical School to UNICOM

TITLE: Evaluating the interoperability of two standard terminologies for dosage form: RxNorm from the National Library of Medicine for the United States and EDQM from the European Directorate for the Quality in Medicines and Healthcare for Europe.

September 15, 2021

Contents

INTRODUCTION

Patients need to be able to safely refill prescriptions when in another country due to travel or relocation. The conventions for identifying therapeutic drugs vary greatly between common drug classification systems in different parts of the world, making it difficult for patients and pharmacists to be sure they are getting the correct drug and dose.

The lack of a comprehensive mapping database creates significant obstacles to safe transnational medication prescribing practices and conducting large-scale international clinical trials. The increasing volume of medications on the market internationally, along with the expansion in global medical tourism[1] and immigration, escalates the need for global infrastructures for drug interaction and clinical decision support to reduce medication errors and protect patient safety. A directive issued in the European Union in 2011 mandated cross-national recognition of prescriptions and provided measures to facilitate health professionals verification prescriptions issued by other member states[2]. While this directive requires that prescriptions be written using a "common name[2]," there are significant variations in medication names and formulations between countries. A global mapping system of pharmaceutical equivalents does not currently exist. Accordingly, a survey of 3,307 German citizens who have prescribed medication while in another EU member state revealed that 12% experienced medication problems related to being prescribed medications different from their normal prescriptions, and 80% reported a lack of health information exchange between countries.[3]

Standardizing the naming system of medications would facilitate developing an international drug information database supportive of clinical decision support aiding clinicians in efficiently determining medication equivalency and interactions. Creating a drug database capable of integrating drug information across countries will support greater interoperability of prescription information, facilitate improved clinical decision support, improve the safety of cross-border medication dispensing, and expand the accessibility of health data for research efforts.

To establish this global drug database, a standardized drug identification system must be established. The Identification of Medicinal Products (IDMP) standards was created in 2012 by the International Organization for Standards (ISO) to outline key characteristics of medicinal product classification systems for international harmonization[4]. One of the key components of IDMP is ISO 11616, which defines the elements required to construct universal Pharmaceutical Product Identifiers (PhPID). PhPIDs are internationally recognizable identifiers for each drug product. They are composed of three main attributes: the active pharmaceutical substance, the dosage form, and the strength,[5] allowing pharmaceutically equivalent drug products to be recognized as synonymous regardless of regional differences in branding and packaging.

To enhance prescribing interoperability across Europe and better adhere to IDMP standards, the European Union commissioned a project called Up-Scaling the Global Univocal Identification of Medicines (UNICOM) to establish a unified drug database for all medications prescribed across the EU and internationally[6]. As the representation of medication dosage forms varies greatly between regional terminologies, a significant barrier to the generation of universal PhPIDs has been the absence of mappings between dosage form representations. The European Directorate for the Quality of Medicines and Healthcare (EDQM) has created a controlled vocabulary relating dose form descriptions to key dosage form characteristics[7], upon which the UNICOM project has structured their dosage form descriptions.

To facilitate harmonization between European and United States drug databases, the World Health Organization, United States Food and Drug Administration (FDA), Uppsala Monitoring Centre, and UNICOM completed a pilot project seeking to align dosage form descriptions between the FDA's Terminology for Structured Product Labeling (SPL) and EDQM[8]. The pilot results were discussed in a June 11th, 2021 webcast and identified difficulties in harmonization between FDA SPL, EDQM, and ISO standards, indicating the necessity for ISO standard revision[8].

The drug terminology system operated by the United States National Library of Medicine is known as RxNorm and is of great significance within the United States and abroad. RxNorm terminology is the basis

for the U.S. Department of Veterans Affairs National Drug File-Reference Terminology (NDF-RT[9]), for the Drug Ontology (DrOn)[10], and for the Observational and Medical Outcomes Partnerships (OMOP) open-source common data model, which is used ubiquitously across the United States and Europe to conduct multinational drug studies[11].

Interoperability between the Canadian drug ontology OCRx and RxNorm has also been demonstrated, further contributing to the international significance of RxNorm serving as a standard drug terminology. While studies have been conducted to evaluate interoperability between the FDA SPL and SNOMED to EDQM[8], there have been no studies to our knowledge that have evaluated interoperability of dosage form descriptions between RxNorm and EDQM. Given the tremendous national and international significance of RxNorm, the objective of this paper is to assess a mapping of RxNorm dosage forms to EDQM-based descriptors.

METHODOLOGY

A description of the dosage form representations in RxNorm and EDQM will be provided, followed by a proposal of how the RxNorm and EDQM data models could be used to generate universal pharmaceutical product identifiers (PhPID). A description of the derivation of EDQM characteristics to the RxNorm dosage forms will be provided. Using the attributed EDQM characteristics to RxNorm dose forms, each RxNorm Dose Form will be a collection of EDQM dose Forms with an identical (or) similar combination of characteristics, based on a simple ontology, developed in the UNICOM Project. This ontology has two levels: a first level based on the ISI characteristic; and a second level based on functional grouping of identical combinations of characteristics, taking into account differences in representing strength.
This ontology was operationalized in Web Protégé. In one application, all the EDQM dose forms were integrated into the ontology, and in another application, all the RxNorm Dose forms.

RESULTS

RXNORM Dose Representation

RxNorm is a normalized naming system established by the United States National Library of medicine for branded and generic pharmaceutical products. RxNorm was created to support interoperability between medical-related terminologies and related knowledge bases across medical applications used in the United States[12]. RxNorm standardizes representation of pharmaceutical ingredients, strength, dose form, and brand name information.

For each unique drug product, an array of codes is generated to represent different identifiers of that drug product. Codes of interest to this investigation include the Ingredient (IN) code derived from the United States Adopted Name (USAN), the Dose Form (DF) code selected from a controlled list of dose forms provided by RxNorm, and the Semantic Clinical Drug (SCD) code, which aggregates descriptions of the product's ingredient, strength, and dose form. RxNorm dosage forms are aggregated into Dose Form Groups (DFG) based on the route of administration, release characteristics or product type. Each dose form is attributed to at least one Dose Form Group, with dose forms often belonging to several Dose Form Groups. For this reason, RxNorm does not provide an ontology for classifying dosage forms but instead a list of defined dosage forms contained in overlapping groups. A sample from the RxNorm representation of dosage forms for sublingual tablets is as follows, illustrating the redundancies of dosage form representations between Dosage Form Groups:

Oral Product (Dosage Form Group)
- Sublingual Tablet
- Capsule
- Tablet

Pill
- Sublingual Tablet
- Buccal Tablet
- Chewable Tablet

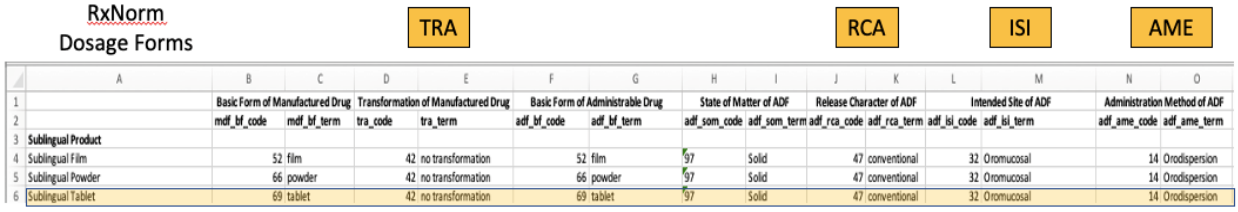
Sublingual Product
- Sublingual Tablet
- Sublingual Film

EDQM Dose Representation

EDQM maintains a set of controlled vocabularies to describe six key characteristics of pharmaceutical dosage forms[7]. These six characteristics include State of Matter, Basic Dose Form, Transformation (TRA), Release Characteristics (RCA), Intended Site (ISI), and Administration Method (AME). Lists of defined terms that characterize dosage forms in each of these six areas are maintained by EDQM. The Basic Dose Form and the State of Matter refer to the drug's form (such as a cream, tablet, implant) and the associated physical state of matter (solid, semi-solid, liquid, or gas). Transformation refers to whether the product requires alteration before administration, such as through dilution or reconstitution. Release Characteristic refers to any alteration of the drug release timing (such as prolonged or delayed-release), Intended Site refers to the anatomical site of drug administration (such as oral, parenteral, or ocular), and Administration Method refers to the method of drug administration (such as via swallowing, chewing, or inhalation). A sample from the EDQM representation of dosage forms for sublingual tablets is as follows:

Oral (Site of Administration)
- Sublingual Tablet
  - State of Matter: Solid
  - Basic Dose Form: Tablet
  - Transformation: No Transformation
  - Release Characteristics: Conventional
  - Intended Site: Oromucosal
  - Administration Method: Orodispersion

MAPPING RXNORM DOSE FORMS TO EDQM DESCRIPTORS



	Basic Form of Manufactured Drug		Transformation of Manufactured Drug		Basic Form of Administrable Drug		State of Matter of ADF		Release Character of ADF		Intended Site of ADF		Administration Method of ADF	
	mdf_bf_code	mdf_bf_term	tra_code	tra_term	adf_bf_code	adf_bf_term	adf_som_code	adf_som_term	adf_rca_code	adf_rca_term	adf_isi_code	adf_isi_term	adf_ame_code	adf_ame_term
Sublingual Product														
Sublingual Film	52	film	42	no transformation	52	film	97	Solid	47	conventional	32	Oromucosal	14	Orodispersion
Sublingual Powder	66	powder	42	no transformation	66	powder	97	Solid	47	conventional	32	Oromucosal	14	Orodispersion
Sublingual Tablet	69	tablet	42	no transformation	69	tablet	97	Solid	47	conventional	32	Oromucosal	14	Orodispersion

Figure 1:Mapping of RxNorm Dosage Forms to EDQM dosage form descriptors. EDQM definitions for drug basic form, transformation, state of matter, release characteristics, intended site, and administration method were used to assign EDQM terms to each RxNorm dosage f

A list of all dose forms recognized by RxNorm was downloaded from Appendix 2 of the RxNorm Technical Documentation (version reviewed July 6, 2020) located on the United States National Library of Medicine website (https://www.nlm.nih.gov). This list of dose forms was uploaded into a Microsoft Excel spreadsheet, and columns were created representing the six characteristics to describe pharmaceutical dose forms under EDQM: state of matter, basic dose form, release characteristics, intended site, transformation, and administration method (Figure 1).

The definition for each RxNorm dose form provided by Appendix 2 of the RxNorm Technical Documentation was used to assign descriptive EDQM characteristics for each dosage form. The EDQM Standard Terms and Internal Controlled Vocabularies for Pharmaceutical Dose Forms (Version 1.2.0) were consulted to manually fill the six characteristics for each RxNorm dosage form based on the provided definitions for each characteristic. To accommodate for the lack of delineation between manufactured dose forms and administrable dose forms, columns were made to describe the state of matter and basic dose form of the drug as it was supplied by the manufacturer (known as the Manufactured Dose Form) and after any indicated transformation (known as the Administrable Dose Form). For dose forms not requiring

RxNorm Dosage Forms | TRA | RCA | ISI | AME

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		Basic Form of Manufactured Drug		Transformation of Manufactured Drug		Basic Form of Administrable Drug		State of Matter of ADF		Release Character of ADF		Intended Site of ADF		Administration Method of ADF	
2		mdf_bf_code	mdf_bf_term	tra_code	tra_term	adf_bf_code	adf_bf_term	adf_som_code	adf_som_term	adf_rca_code	adf_rca_term	adf_isi_code	adf_isi_term	adf_ame_code	adf_ame_term
3	**Sublingual Product**														
4	Sublingual Film	52	film	42	no transformation	52	film	97	Solid	47	conventional	32	Oromucosal	14	Orodispersion
5	Sublingual Powder	66	powder	42	no transformation	66	powder	97	Solid	47	conventional	32	Oromucosal	14	Orodispersion
6	Sublingual Tablet	69	tablet	42	no transformation	69	tablet	97	Solid	47	conventional	32	Oromucosal	14	Orodispersion

2

transformation, the Manufactured Dose Form and the Administrable Dose Form state of matter and basic dose forms were equivalent. The release characteristics, intended site, and administration method were subsequently assigned in reference to the administrable dose form.

PROPOSED INTEGRATION OF RXNORM AND EDQM

We propose applying the EDQM framework of drug dosage form descriptions to RxNorm dosage forms. For each drug product in RxNorm, the DF code can be mapped to RCA, ISI, TRA, and AME codes. The mapping of RxNorm dosage forms to these characteristics is described in the previous section. Additionally, the IN and SCD codes can be mapped to globally accepted ingredient and strength codes. Accordingly, the decision of which ingredient and strength codes should be used universally for this purpose is outside the scope of this paper. The combination of ingredient, strength, and four dosage form descriptor codes can then be used to generate a universal pharmaceutical product identifier (PhPID) to identify pharmaceutical products independent of regional naming conventions (Figure 2).

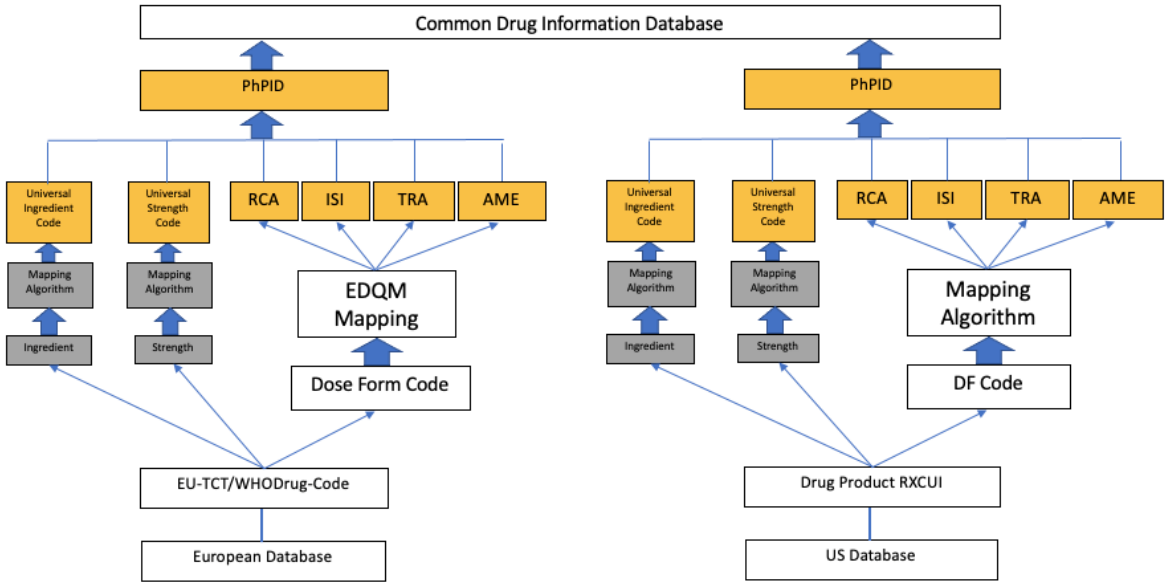


Figure 3: The proposed common data model for generating pharmaceutical product identifiers from EDQM

and RxNorm. The mapping algorithm between the RxNorm DF code and the release characteristics (RCA), intended site (ISI), transformation (TRA), and administration method (AME) are defined in this paper.

COMPARING RxNORM AND EDQM WITH A COMMON ONTOLOGY

An ontology was created based on the EDQM site of administration, with the site of administration as the primary grouping mechanism. Each RxNorm dosage form was fitted into this EDQM-based dosage form ontology, with preservation of the transformation, release, administration method, and intended site of administration codes previously derived. This ontology was uploaded to the online ontology-builder Webprotege and is accessible here: https://webprotege.stanford.edu/#projects/34af25bd-e27b-4b36-bbc1-4498b0706971/edit/Classes?selection=Class(%3Chttp://unicom-project.eu/doseforms.owl%23DoseForm%3E).

give a screenshot splitscreen of the ontology opened on the dose form group in both terminologies.

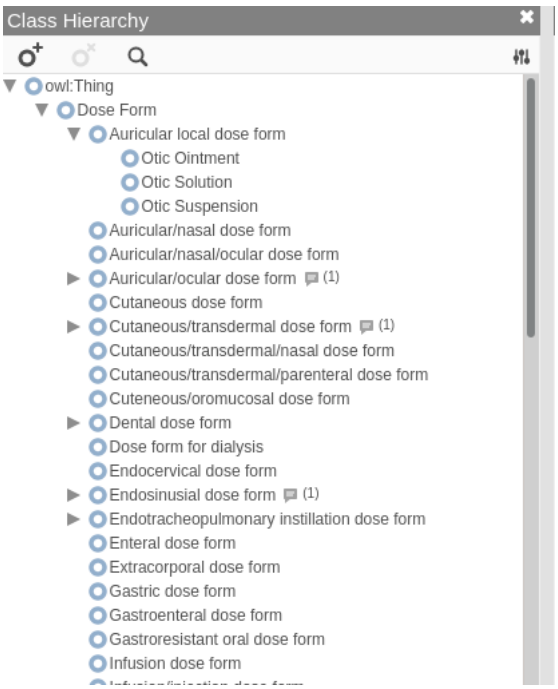


Figure 4: RxNorm anchored in the dose form groups

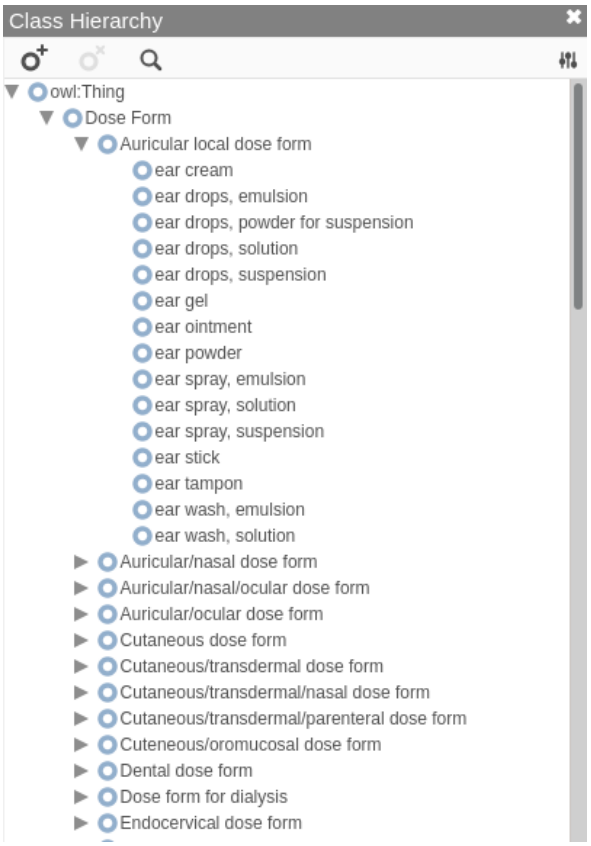


Figure 5: EDQM anchored in the dose form groups

DISCUSSION

The mapping of RxNorm dose forms to EDQM descriptors revealed several issues requiring attention before reliable mapping is established. There is a significant disparity in granularity of dosage form descriptions between RxNorm and EDQM, with RxNorm recognizing 179 dose forms and EDQM recognizing 428 dose forms (not including the veterinary dose forms). While there is considerable overlap between the dosage forms in both systems, some dosage forms are exclusive to RxNorm (such as urethral suppository) and many exclusive to EDQM. RxNorm Dose Form Groups vary significantly in the rationale behind granularity and grouping, resulting in Dose Forms often belonging to several Dose Form Groups. Due to the inconsistent granularity of RxNorm dosage form groups and redundant dose form

representations across groups, it would be recommended that ontologies and data models founded on RxNorm be based on the list of dose forms instead of the dose form groups. This reflects the structure of OMOP, which omits the use of dosage form groups[13].

The data quality issues and redundancies in dose form representations discussed here have also been described as challenges to using RxNorm for clinical decision support[14]. RxNorm lacks the attributes of a traditional ontology required to facilitate universal interpretability, further supporting our proposal to describe RxNorm dose forms with the more structured EDQM or to use the ontology created by mapping RxNorm dose forms to EDQM.

The mapping of RxNorm dosage forms to EDQM dosage forms encountered many issues described by the FDA pilot project, namely differentiation between the manufactured and administrable (post-transformation) dosage forms[8]. Accordingly, RxNorm would benefit from alignment with IDMP/ISO standards, as recommended by the WHO_UMC/FDA pilot project.

Additionally, the assignment of EDQM characteristics to dosage forms must be standardized across countries to prevent the creation of redundant PhPID due to inconsistent assignment of characteristics. This point was also emphasized during a UNICOM Community of Experts Webinar during discussion of the varying assignment of the basic dose form characteristics dispersion, suspension, and solution to the Pfizer coronavirus vaccine by different nations, leading to the generation of three unique PhPID for the same product[15].

CONCLUSION

Lacking international standardization of basic dose forms and other key characteristics to pharmaceutical products will lead to the inefficient generation of pharmaceutical product redundancies and the omission of regional pharmaceutical products from global PhPID catalogues. Overall, the collaboration between RxNorm and international harmonizing agencies is essential to maintaining adherence to IDMP standards and facilitating the systematic generation of PhPID. The mapping of RxNorm to EDQM dosage form characteristics shows promise in harmonizing drug descriptions internationally and supporting important applications such as OMOP, NDF-RT, and DrOn.

REFERENCES:

1. Hong YA. Medical Tourism and Telemedicine: A New Frontier of an Old Business. *J Med Internet Res*. 2016;18(5): e115. doi:10.2196/jmir.5432

2. EU Directive 2011/24/EU on the application of patients' rights in cross-border healthcare: Legislation Guidance for Patient Organisations. :20.

3. Panteli D, Wagner C, Verheyen F, Busse R. Continuity of care in the cross-border context: insights from a survey of German patients treated abroad. *Eur J Public Health*. 2015;25(4):557-563. doi:10.1093/eurpub/cku251

4. Revised IDMP standards to improve description of medicinal products worldwide. ISO. Accessed August 26, 2021. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/news/2017/10/Ref2234.html

5. Research C for DE and. Identification of Medicinal Products (IDMP). *FDA*. Published online August 23, 2019. Accessed August 26, 2021. https://www.fda.gov/industry/fda-resources-data-standards/identification-medicinal-products-idmp

6. UNICOM_Gap-Analysis.draft_.v12-1.pdf. Accessed August 25, 2021. https://unicom-project.eu/wp-content/uploads/2020/09/UNICOM_Gap-Analysis.draft_.v12-1.pdf

7. Standard Terms Database | EDQM - European Directorate for the Quality of Medicines. Accessed August 26, 2021. https://www.edqm.eu/en/standard-terms-database

8. Identification of Medicinal Products: Path to Global Implementation - 06/11/2021 - 06/11/2021. FDA. Published June 16, 2021. Accessed August 26, 2021. https://www.fda.gov/drugs/news-events-human-drugs/identification-medicinal-products-path-global-implementation-06112021-06112021

9. Zhu Q, Jiang G, Chute CG. Profiling structured product labeling with NDF-RT and RxNorm. *J Biomed Semant*. 2012;3(1):16. doi:10.1186/2041-1480-3-16

10. Hanna J, Joseph E, Brochhausen M, Hogan WR. Building a drug ontology based on RxNorm and other sources. *J Biomed Semant*. 2013;4(1):44. doi:10.1186/2041-1480-4-44

11. Kent S, Burn E, Dawoud D, et al. Common Problems, Common Data Model Solutions: Evidence Generation for Health Technology Assessment. *Pharmacoeconomics*. 2021;39(3):275-285. doi:10.1007/s40273-020-00981-9

12. RxNorm. Accessed August 26, 2021. https://www.nlm.nih.gov/research/umls/rxnorm/index.html

13. implementation_international_drug_vocabulary [Observational Health Data Sciences and Informatics]. Accessed August 26, 2021. https://www.ohdsi.org/web/wiki/doku.php?id=implementation_international_drug_vocabulary

14. Freimuth RR, Wix K, Zhu Q, Siska M, Chute CG. Evaluation of RxNorm for Medication Clinical Decision Support. *AMIA Annu Symp Proc*. 2014;2014:554-563.

15. UNICOM. *UNICOM - Community of Expertise: Representation of Clinical Information*.; 2021. Accessed September 20, 2021. https://www.youtube.com/watch?v=EpWKdw6uelA

Document version	D8.1_ANNEX 6_Procedures_Production¨_PhPID
Authors	Malin Fladvad, WHO-UMC, UNICOM-WP1 Anna Baumgarten, WHO-UMC Robert Vander Stichele, UNICOM-WP8
Changes	Version of March 2021, to be completed with the results of the WHO_UMC/FDA Pilot Report

Procedures for the production of global identification numbers for Pharmaceutical Products (PHPID_SUB_L4)

Version 1.2

Date : March 12, 2021

Version History

Preliminary Note	August 6, 2020	Robert Vander Stichele	Initial draft
Procedures Version 1.1	September 15, 2020	Malin Fladvad Anna Baumgarten Robert Vander Stichele	After comments of Julie James, Leonora Grandia, and discussions with RVS, MF, AB
Procedures Version 1.2	March 12, 2021	Malin Fladvad Anna Baumgarten	Adding conclusions from pilot of 6 substances

Content

Introduction

The ISO/CEN standardization bodies are engaged in the development of a global system for identification of medicinal products since more than a decade. The original focus was to support worldwide pharmacovigilance and included engagement of the pharmaceutical industry; the ICH collaboration between the regulatory agencies of EU, USA and Japan; and the WHO.

A collection of ISO/CEN Standards was developed under the IDMP heading, governing the identification of substances, dosage forms, measurement units

The idea was to harmonize description of medicinal products in different jurisdictions by implementing a global system of identification at 3 level:

- Pharmaceutical Product (PHPID): a global identifier for the abstract description of a medicinal product, independent of the jurisdiction and the company.
- Medicinal Product (MPID): a global identifier for a medicinal product authorized within a specific jurisdiction, and marketed by a specific pharmaceutical company
- Medicinal Product Package (PCID): a global identifier of a medicinal product package as marketed by a specific pharmaceutical company within a specific jurisdiction, with a defined pack size

The construction of the PHPID should be governed at the global level, based on a MD5 Hash Function, taking into consideration numerical representations of substance, dosage form and strength and the internal order of the fields.

The construction of MPID and PCID would be governed at the supranational and national level, using the global PHPID, completed with a set of standardized characteristics, taken from the IDMP standards.

The EU supported preliminary work and potential implementation of these standards, through the research projects EPSOS, OpenMedicine, and now UNICOM.

Objective

Within UNICOM, the community of Standard Development Organizations (SD0s), coordinated in Work Package 1, contacted the WHO UMC in Uppsala to propose a procedure for the production of PHIPID identifiers.

This proposal would be elaborated using on an exemplary basis the work on the UNICOM Pilot Product List, providing a sample of more than 30 substances and their medicinal products, as marketed in 4 European countries.

The joint WHO / UNICOM proposal could be developed by the first half of 2021, and then be submitted to a wider consultation round.

The proposal should include:

- the procedures of choice for the 3 basic identifying concepts: substance, dosage form, strength (unit and value),
- the procedures of choice for the numerical representations for these concepts,
- the procedure for presentation of these numerical representations to the chosen HASH function,
- the procedure for making the resulting PHPID publicly available,
- a quality assurance process.

Project Scope

According to the ISO standard for Pharmaceutical product, ISO 11116, PhPID shall be presented for both active substance and specified substance, each containing four PhPID identification levels. The PhPID shall be generated using the corresponding ISO standards:

Substance ISO 11238 and ISO/TS 20440

Administrable dose form (ISO 11239)

Units of measure (ISO 11240)

Table 1 — Four levels of PhPID

PhPID active substance stratum	PhPID_SUB_L1 → Substance(s) PhPID_SUB_L2 → Substance(s) + strength + reference strength PhPID_SUB_L3 → Substance(s) + administrable dose form PhPID_SUB_L4 → Substance(s) + strength + reference strength + administrable dose form
PhPID specified substance stratum	PhPID_SpSUB_L1 → Specified substance(s) PhPID_SpSUB_L2 → Specified substance(s) + strength + reference strength PhPID_SpSUB_L3 → Specified substance(s) + administrable dose form PhPID_SpSUB_L4 → Specified substance(s) + strength + reference strength + administrable dose form

This project scope will mainly explore PhPID for active substance on the fourth level, calculating PhPID_SUB_L4

The procedure of choice for the 3 basic Identifying concepts

Substance

The first task is to identify the active ingredient with therapeutic intent at the most appropriate level of granularity. For Chemicals that means that salts/esters of the substance must be defined, if relevant. The option chosen here is to define the PHPID_SUB_L1 at the level the moiety with the specific salt or ester, if relevant. If the active ingredient is a salt or ester, the reference substance needs to be identified and expressed based on anhydrous free acid, anhydrous free base or a substance created to express activity, ie active moiety, The level of granularity needed for unique identification of a substance will be based on the current investigation by ISO WG6.

The substance data used for the project is based on UNICOM Pilot product list where the following details for each substance is captured:

For each substance at the chosen granularity level, e.g. SIMVASTATIN ACID (unii=9L6M5TH46B) and SIMVASTATIN (UNII=AGG2FN16E), the (molecular) weight of the moiety and of the combination of moiety with salt and ester should be made explicit. The type of substance should also be systematically defined (chemical, mineral, protein, …).

The substance should be present in at least one authorized medicinal product in at least one jurisdiction or an internationally recognized experimental product.

For each substance at the chose granularity level, the identification numbers in different systems should be identified and listed in a mapping file (WHO Drug Dictionary, INN Modified, SNOMED, UNII, CAS, EU-SRS, SPOR, US-SRS, …).

The identification number of the substance to be submitted to the HASH function must be chosen and the order of the substances if >1. An investigation of the requirements for the identification number will be conducted and a proposal will be made by WHO UMC.

There are several issues regarding substances that needs clarification:

- Should the reference substance only be expressed when the active ingredient is identified as a salt or ester?
- For some substances (e.g. biologicals) further specifications are possible, but do not necessarily need to be taken into account for the PHPID production.
- If a substance belongs to a collection of substances with the same therapeutic moiety, all relevant salts and esters in that collection should also be defined and submitted to the above procedures. It should also be established how to refer to INN in for the salt and ester variations[1].
- In Adverse Event Reports it is possible that brand names or unspecified substances are mentioned. Hence, it may be necessary to include in the collection of possible specific substances class of "substance unspecified". It is hoped that the need for this will diminish, as Companies and Marketing authorization Authorities adopt IDMP in the future. These unspecified substances should however not be used to generate PhPID calculations on level 2-4.

Dosage Form

EDQM is the system of reference for dosage form that follow ISO 11239, is recognized by EMA, and will therefore be used to express dosage forms in this project, also for numerical presentation. It is proposed to choose the most granular administrable dose form for PHPID production, not the basic dose form.

For those dosage forms that do not undergo transformation processes from distribution to administration phase the pharmaceutical dose form will be used if an administrable dose form is not available within EDQM.

Some issues related to dose forms needs further elaboration:

- When dose forms are expressed differently within different jurisdictions as seen for some of the Covid-19 vaccines in table 1, a clear base for decision of which dose form should be used to create a global PHPID is necessary.

Authority of approval	Administrable dose form
EMA	dispersion for injection
FDA	suspension for injection
UK	solution for injection

Table 1. Example of dose forms assigned for the Covid-19 vaccine Comirnaty within different authorities

[1] An INN is usually designated for the active part of the molecule only, to avoid the multiplication of entries in cases where several salts, esters, etc. are actually used. In such cases, the user of the INN (pharmacopoeia commissions, regulatory bodies, pharmaceutical manufacturers) has to create a modified INN (INNM) himself; mepyramine maleate (a salt of mepyramine with maleic acid) is an example of an INNM. When the creation of an INNM would require the use of a long or inconvenient name for the radical part of the INNM, the INN programme will select a short name for such a radical (for example, mesilate for methanesulfonate).

- If the administrable dose form can be expressed in differently in ml, but the posology in drops, the relationship between these two forms of expression should be made explicit (eg. 20 drops per ml).
- For solid oral forms, galenic features of controlled release are important sometimes be specified (rapid release, prolonged release BID or OD), but may share the same strength.

To summarize, the ontology of dosage forms in EDQM should be carefully studied to understand the implications with regard to for example local or systemic action, and to traditional gastro-intestinal absorption and avoidance of the liver circulation, also in relation to the characteristics of EDQM. The investigation should lead to suggestions for improvements to EDQM.

The new alternative proposal that is under investigation in ISO workgroup 6, using the characteristics of the PDF instead of the described system that maps to EDQM, is not in scope of this project but will be further investigated within ISO TC 215 workingroup 6.

Strength

For the unit of strength, the UCUM standard should be used and the numerical values of UCUM should be expressed for the following elements of strength has associated with the substance or specified substance:

- The presentation strength, also called basis of strength, is the strength of a substance described as a qualitative term describing the discrete unit in which a Pharmaceutical Product is presented, such as weight per tablet.
- The concentration strength is the strength of a substance expressed as the amount of substance per unit of measurement, such as millilitre or gram
- The reference strength shall be expressed based on anhydrous free acid, anhydrous free base or a substance created to express activity, ie active moiety

Rules and roles should be made explicit to express the above concepts for strength and their extent of integration in the PHPID. If the reference substance is identified as an active moiety, will the reference strength be required?

If there is no exact strength related to the substance, the strength interval will be expressed as RTO<PQ,PQ>data type according to ISO11616.

There are several issues regarding strength that needs to be investigated:

- The denominator should be made explicit and can be of a different nature for oral solid forms and liquids (100mg per tablet for oral forms, 250mg/ml for syrups, 1000mg per sack for soluble powder).
- For liquids, the strength shall be expressed per total volume per container and strength (concentration) per unit volume. The strength concentration per unit volume shall be calculated from the strength per total volume of the container. There is a difference between weight per volume and concentration (e.g 5mg/2ml vial) or 2.5mg/ml in a 2ml vial). (in case of concentration the value of the denominator is an implicit 1)[2].
- For injectables, the difference between unidose vials and multidose vials has numerous implications for expression of strength, pack size and units of administration in the signatura.
- For a patch, strength shall be expressed as per time unit or per each patch according to approval. If not rate, use quantity per each/contained by each.
- EDQM has ‘transdermal patch’ as dose form for a patch with systemic action and ‘cutaneous patch’ is meant for a local effect. The difference between the site of action (systemic or local) should be expressed as such, not ‘hidden’ in the way of expression of strength.

[2]For PhPID and liquid preparations, the strength shall be expressed per total volume per container and strength (concentration) per unit volume, at every instance of PhPID level 2 and 4.

- For topical products (creams, ointments), the establishment of the unit of administration in the posology is not as self-evident as it is in solid oral dosage forms.

To determine the nature the strength expressions and the above-described issues for different types of product, the EU IG described patterns for expressions of Pharmaceutical Product, will be evaluated[3]. Using the EU IG patters would generate data for PHPID generation for Covid-19 vaccines according to table 2.

Product	Substance	Strength by presentation	Strength by concentration	Administrable dose form
Covid-19 vaccine Pfizer (Comirnaty)	Tozinameran	NA	100 µg/ml	Dispersion for injection
Covid-19 vaccine Moderna	COVID-19 vaccine mRNA (mRNA 1273)	NA	200 µg/ml	Dispersion for injection
Covid-19 vaccine AstraZeneca	COVID-19 vaccine NRVV Ad (ChAdOx1 nCoV-19)	NA	2.5×10^8 infectious units	Suspension for injection

Table 2. Using pattern 3A from the EU IG, strength presentation would not be applicable for PhPID generation, only strength concentration.

There are also some discrepancies seen for which units are used to express strength within different jurisdictions for the Covid-19 vaccines, see table 3. Like the dose forms, a clear base for decision of expression of strength for Global PhPIDs are necessary.

Authority of approval	Strength per dose (0.5 ml)
EMA	2.5×10^8 infectious units
UK	5×10^{10} viral particles
Australia	5×10^{10} viral particles

Table 3. Expression of strength for COvid-19 vaccine from AstraZeneca.

The procedures of choice for the numerical representations

Substance

A choice needs to be made between the available candidates; EUTCT, UNII, UMC Substance ID etc. This choice should be acceptable at an international and global level and should fulfil the requirements in the ISO/IEC 15459, specifying the common rules applicable for unique identification.

Any chosen system must be able to reflect the chosen level of granularity for substances.

Dosage form

The EDQM number will be used for this pilot.

[3] https://www.ema.europa.eu/en/human-regulatory/research-development/data-medicines-iso-idmp-standards/spor-master-data/substance-product-data-management-services#eu-idmp-implementation-guide---version-2.0-section

Strength

Value+UCUM number should be used. Since the UCUM nomenclature do not provide a list with fixed values, the UMC value list will be used for this project. There should also be a consensus around how strengths are expressed,for example 0.1 g or 100 mg.

Example of numerical representations

Product	Substance ID	Strength by presentation	Strength by concentration	Administrable dose form
Covid-19 vaccine Pfizer (Comirnaty)	36667	NA	100 µg/ml	50077000
Covid-19 vaccine Moderna	35854	NA	200 µg/ml	50077000
Covid-19 vaccine AstraZeneca	35853	NA	2.5×10^8 infectious units	11202000

Table 4. Numerical representation of 3 Covid -19 vaccine using the UMC simplified substance ID and EDQM dose form code. The reference substance and reference strength is identical to substance and strength in this case and have been excluded from the table.

The procedure for presentation to the chosen MD5 HASH function

Order of basic characteristics in single products

There is a choice to be made between two possible sequences, when substance comes first:

- Substance / dose form / strength
- Substance / strength / dose form

UMC have used the order of substance/ strength/ dose form for this pilot to keep the substance information with its relating strength but have not made any further investigations in relation to other conceptual systems.

Combination products (combination of substances)

A proposal needs to be made for the order of substances, in case of combination products and how their respective strength will be presented, after each substance or gathered consecutively?

There is also a decision to be made for other cases:

- Complex medicinal products with many substances (e.g. multivitamins). How and when is a PhPID useful and possible to create in a harmonized way (since countries sometimes define the number of active ingredients differently for these product)?
- The policy for outdated FDC (fixed dose combinations) e.g. combinations of antipyretics, cough products and antibiotics.
- What will the policy be for border line cases (adjuvants[4])
- Are complex parenteral hospital products for infusion in scope?

Multiple products packaged as a kit with intent to being administered as one medical product shall be assigned one overarching PhPID according to ISO.

[4]ISO 11616: If an adjuvant is applicable for eg vaccine, the adjuvant term and term ID shall be displyed with the active substance (s) and specified substance(s) terms for the product on all applicable PhPID levels. This association shall be made by directly associating the assigned PhPID to MPID and PCID.

Hybrid products

Both combinations of pharmaceutical products[5] and combinations with medical devices or diagnostics[6] needs to be considered.

The hash function

A few different hash functions were investigated:

- MD5 (128 bit)
  - pros: fast
  - cons: considered broken, known to have hash collision
- SHA2 (256 bit and 512 bit)
  - pros: considered secure (have vulnerabilities but are not considered to be too serious)
  - cons: slower then MD5
- SHA3
  - It is yet to gain widespread support and implementation and will be unlikely to do so until significant flaws in SHA2 are found that necessitate an update.

The MD5 algorithm is a much faster hashing algorithm then SHA2 but it is not cryptographically secure. Its main application is data integrity verification. It is possible to force a collision between two MD5 hashes if you control the input of both hashes.

The procedure for making the resulting PHPID publicly available

The procedures for producing the PHPID (the current document) should be made publicly available as a versioned, living document. There should be a publicly available Linked Open Data PHPID database with:

- the PHPIDs
- the basic concepts and their numerical representations (enabling checks of Hash functions)

Additional linked data may be beneficial (for example other identifiers (e.g. INN, UNII, SPOR, EU-SRS, CAS, SNOMED and the link to ATC/ROA/DDD methodology)

Other instances could govern links to standardized indications, contra-indications, side-effects, etc. but also to drug classifications (multi-axial SNOMED drug classes, WHO Standardized Drug Groupings, table of content of medicinal product dictionaries, simplified classifications for patients and medical students

[5] From EU Implementation guide v 2.0: A medicinal product may contain one or more "pharmaceutical product(s)" (e.g. a kit containing vaginal tablets 500 mg and a vaginal cream 10% or a kit containing a combination of norethindrone acetate and ethinyl estradiol tablets and ferrous fumarate tablets). In these instances, a pharmaceutical product section is to be completed for each "pharmaceutical product".

[6] From EU implementation guide v 2.0. Where applicable, the technical concept of a "pharmaceutical product" can also include information on a medical device if it is an "integral part" of the medicinal product and supports the pharmacological/metabolic/immunological action of the medicinal product, for example the scaffolding or net for a cell therapy medicinal product in accordance with Regulation (EC) No 1394/2007. Any other device co-packaged (e.g. spoon, syringe) or integral (e.g. pre-filled pen) with the medicinal product must be recorded as part of the packaged medicinal product. Strength is not applicable for devices

The quality assurance process.

The following should be investigated for the quality assurance process:

- WHO UMC would be the executive responsible organization.
- A suitable business and funding plan should be made
- Appeal and feedback procedures should be in place
- A multi-stakeholder steering committee should govern a quality assurance process.

In a first phase, the procedures could be tested on the UNICOM Pilot Product List, and submitted to a round of comments inside UNICOM.

Ultimately it is for EMA, FDA and other regulators to decide on the suitability of this approach.