OMS Data Quality Standard

Webinar for OMS users, 25 September 2018

SUBSTANCE PRODUCT REFERENTIALS ORGANISATION

European Medicines Agency

Welcome to the OMS Data Quality Standard webinar

- Duration - 2 hours

- Webinar is in listen only mode

- Participants can type questions in the chat box

- We will respond to questions at the end of the session

OMS data quality standard

1. General rules

- Data in Latin characters
- Versioning
- IDs management

2. Organisation name & type

- Preferred main name
- Alternative names
- Organisation type

3. Location details

- Main address
- Localised address
- Communication details

4. Reference IDs

- Mergers
- Splits
- Mappings

5. Quality improvement process

- Improvement process
- Key user group

Data in Latin Characters

- OMS will support EN alphabet letters & all other non EN (EEA official language) alphabet letters

  - <u>Org. names:</u> "Blueclinical Investigação E Desenvolvimento em Saúde Lda.", "Svenska Örtmedicinska Institutet AB", "Arrow Läkemedel AB", etc.

  - <u>Loc. addresses:</u> "Itäinen Pitkäkatu 4b", "Bøge Alle 6/8", "Dronninggårds Alle 136", etc.

- Non Latin Characters i.e. Greek and Cyrillic can be stored in OMS. However, OMS services cannot validate the data

  - <u>Org. alternative names:</u> "Φαρμασέρβ-Λίλλυ Α.Ε.Β.Ε.", "Сервие България ЕООД", etc.

  - <u>Loc. Address:</u> "Бул.янко Сакързов 26", "Αγίου Δημητρίου 63", etc.

Versioning of Organisation & Location data

- Versioning is based on time stamps (Active Start Date vs. Active End Date)

- Organisation and Location entities have their own time stamps (independent from the other)

- The history of the Organisation and Location records start when these entities are first mastered (ORG-ID or LOC-ID generated)

- Changes to legacy data (e.g. xEVMPD – Art.57) which happened before the data was first loaded to OMS are not reflected in the OMS data versions

- Any change to the Organisation and Location entities' or child entities' related data create a new version in OMS. Some meta data changes can trigger a new version too

IDs management

- OMS generates two IDs, Organisation ID (for Organisation entity) and Location ID (for Location entity)

- An Organisation/Location can have multiple OMS IDs. This happens when 2 mastered Organisation/Locations are merged. Merged record keeps all the IDs

- The ID associated to the oldest master Organisation/Location is kept as the main ID and shown on OMS web portal/API. However, all these IDs can be used for searching in OMS

- When an Organisation is split in OMS, i.e. one of its Locations is moved to a new or another existing Organisation, the Location ID is kept after the split

- Merges and Unmerges may change the 'main' ID published in OMS

2. Organisation name & type

PharmaCo

1. General rules

3. Location details

4. Reference IDs

5. Quality improvement process

- Organisation name comprises of '**Name**' + '**Legal entity type**' (not always included)
- **Multiple names** can be stored although only one name is published as the main name
  - **Preferred main name** is in **EN** although it can be in other languages if no EN is available
  - Abbreviations and symbols should be avoided unless they are an integral part of the legal name
  - Other names, which can be in different languages, are stored as '**Alternative names**'

Example

Organisation Name:	Pharmachem Poessneck GmbH & Co. KG
Alternative Name:	DE - Pharmachem Pößneck GmbH & Co. KG

EUROPEAN MEDICINES AGENCY

- **OMS Data Quality standard** provides a guidance on standard legal entity type formats per jurisdiction

- OMS Data stewards follow this guidance when validating names

Country	Reference	Legal Entity naming in organisation names (in EN and National languages)
IE	Irish companies registration office	Limited, Public, EEIG,
IT	Italian Chamber of Commerce	S.r.l., S.p.A., GEIE,
LI	Liechtenstein Enterprise Register	AG, GmbH,
LT	State Enterprise Centre of Registers	SIA, AB, EEIG, UAB,
LU	Luxembourg Business Register	S.à.r.l., GEIE,

- Organisation name acronyms can be recorded when provided only. It must be a short data value (max. 20 characters)

  - It is not validated by EMA Data stewards – stored as provided

- Organisation type is a controlled vocabulary defining the category of the organisation which can be defined by 'type' or by 'size'

  - Two levels of values

  - For values referring to 'type':

    ➢ Level 1 value is mandatory

    ➢ Level 2 value is enriched by

    Data stewards only

Organisation Type:	Industry
	Pharmaceutical company
Organisation Type:	Health care
	Hospital/Clinic/Other health care facility
Organisation Type:	Regulatory Authority
	EEA National Competent authority

  - Values referring to 'size' are not editable – sourced from EMA SME data base

3. Location details

1. General rules
2. Organisation name & type
4. Reference IDs
5. Quality improvement process

Location Details

- Location address is validated by the OMS address verification tool (Address Doctor)

- Main address vs. Address Localised

  - Main address is identified as the address in EN

  - Address Localised refers to address representations in other languages (these are not translations)

    Example

- Address data is validated against National postal services' address reference data

  - Abbreviations are often transformed to full words, address lines may be swapped (line 1 with line 2), street name may be enriched with a number, etc.

  - Localised address is auto generated based on the data in the reference data files

- Address quality is defined by its '**mailability**' once it is verified i.e. likelihood of successful delivery to the validated address. This can vary between jurisdictions

- Main address data business rules:

  - County[1] or State values are enriched even if not provided except for predefined jurisdictions (e.g. France) – *the exceptions can be extended in the future*

  - Sub-locality or District value can be enriched when available in reference address data (see previous slide for an example – *Donaustadt* district of Vienna*)*

  - County value is not enriched if it is the same as city value (e.g. City = 'Barcelona' & County = 'Barcelona', City = 'Vienna' & State = 'Vienna')

Submissions of MA applications with above enriched/transformed data should not be a cause of rejection/invalidation of the application.

[1]Region or Province or Department/Canton is represented by County

Address transformation examples*:

Before address verification	After address verification
Aven. / Av. / Ave	Avenue
St. / St	Street
Indus. / Ind.	Industrial
Blvd / Blvd.	Boulevard
With floor number, department name, …	Without floor number, department name, …
Without street number	With street number
6-8 (street number)	6/8 (street number)
Street name contains the word 'street'	Street name does not contain the word 'street'
Renaissance Walk, n. 5	5 Renaissance Walk

*Examples are in EN but similar transformations/standisation apply to non EN addresses

Location Details – Main address

- Often, provided address is not verifiable when street number is not provided, post code is not recognised, only PO Box provided (no post code), no street names provided, not street number provided, etc.

  - In these cases, GPS coordinates are not generated by the OMS address verification service

- When address reference data is found to be incorrect or inappropriate, the OMS Data steward can ignore the data transformation/enrichment performed by the address validator. **Only done if absolutely necessary**

- If Location address is not in EN when Location is first mastered, OMS may transform the main address into EN characters and generate the address in the non EN characters as localised address
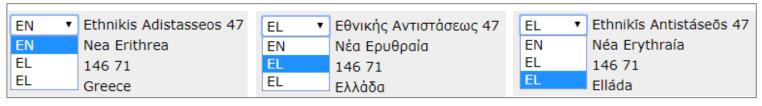
Location Details – Localised address

- Generated by OMS based on address reference files data

- It is the same address but in a non EN language

- Post code and PO Box are not generated – main address contains this data

- It is generated when main address is verified (Good quality). If Address Doctor is ignored when verifying the main address whilst there is a localised address, the changes in main address must be manually reflected in the address localised

- If the address generated does not have non EN characters correct, this is because the relevant postal services provide the address data like that

- When an OMS user submits a change request to add an Organisation or a Location with the location address in non EN, and with non EN characters (ü,í,å,ø,ã,...etc.), they should refer to the non EN address in OMS

- Except for GR and BG, there should be only one localised address for a Location

  - For GR/BG, two localised addresses can be created provided they exist in the address reference files

  - One address can be in Greek/Cyrillic characters, and the second address in Latin characters

Example

Location Details – Communication Details

- Refer to Email address and Telephone number – Not mandatory

- They relate to the generic Location contact details

- **Telephone**:

  - Can only be numbers – 15 character value max

  - Country code must be preceded by '+'

- **Email**:

  - Preferably, a mailbox address - not a personal email address

- Stored only if provided through a Change Request

- Avoid requesting multiple email/telephone numbers per location – only 1 set needed

4. Reference IDs

1. General rules

2. Organisation name & type

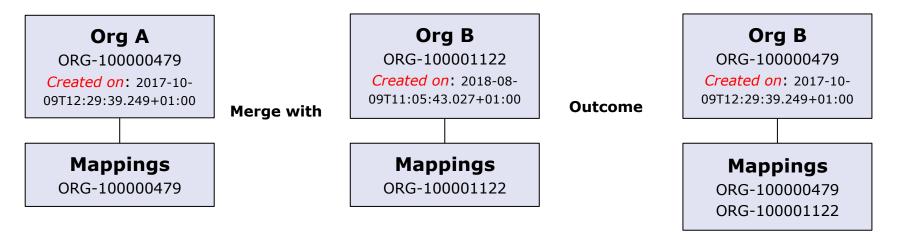3. Location details

5. Quality improvement process

Reference IDs - Merges

- Formats: Organisation → ORG-100000479; Location → LOC-100004173

- **Organisation Merge**

  Business Rule: ID associated to the oldest Organisation (based on creation time stamp version) is published as the 'main ID'
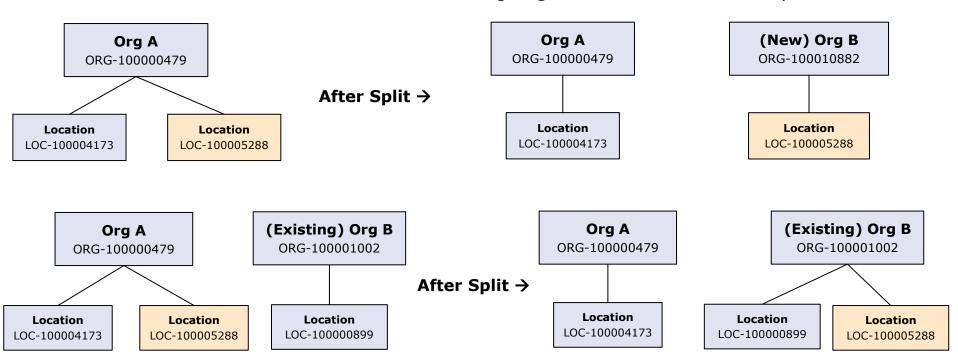
Reference IDs - Splits

- Organisation Split:

  Location is moved to a new or another existing Organisation and its ID is kept

Reference IDs - Mappings

- All OMS and external (to OMS) IDs are stored under 'Mappings'

- Types of IDs published in OMS:
  - Organisation & Location IDs
  - xEVMPD EV_Codes
  - EudraGMDP numbers
  - DUNS numbers
  - GS1 numbers

Location ID:		LOC-100002660
Address:	EN	100-110 Esplanade Du General De Gaulle Coeur Defense Paris La Defense Cedex 92931 France
GPS Location:		48.890610, 2.243290
xEVMPD Code:		ORG4002, ORG3350, ORG1
EudraGMDP Number:		5744, 19332, 41192

Location ID:		LOC-100010911
Address:	EN	Hallerstrasse 7 Bern Be 3000 Switzerland
GPS Location:		46.951940, 7.435120
Location Email:		anfragen@swissmedic.ch
Location Phone:		+41 584620211
DUNS ID:		481761612
GS1 ID:		7611746000006

- Organisation and Location entities have their own separate mappings

- DUNS and GS1 numbers and the only IDs an OMS user can request to add/correct/remove

- All other IDs are either generated by OMS or loaded from EMA legacy systems

- OMS allows to search for any published IDs included in the mappings

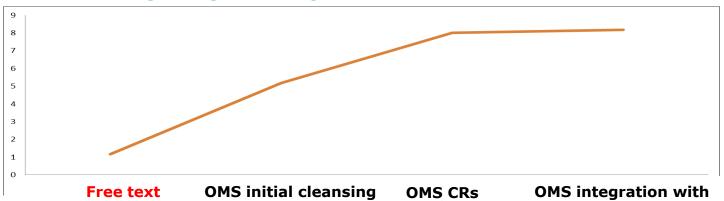# 5. Quality improvement process

1. General rules
2. Organisation name & type
3. Location details
4. Reference IDs

EUROPEAN MEDICINES AGENCY

OMS Data Quality will improve over time

Free text

OMS initial cleansing

OMS CRs

OMS integration with business processes

We started with free text. We experienced a big gap in data quality, which was mainly due to lack of data standards, lack of governance and standardised processes.

We focused on mastering of 'legacy data' from Telematics systems, *i.e.* **xEVMPD – Article 57, EudraGMDP, and 3 other EMA corporate systems**. Data **cleansing, standardisation, consolidation** then publication in the OMS dictionary**.**

OMS change request (CR) helps to improve data quality in OMS. Since December 2017 we received over 1500 OMS CRs.

As SPOR data will be used by business process the quality of data will improve. How? Through OMS CR process (if data needs change or addition).Today OMS supplies master data to: eAF, IRIS, User Registration portal.

EUROPEAN MEDICINES AGENCY

- This group will provide a forum where EMA and RMS/OMS stakeholders can share and exchange experiences, and identify opportunities for improvement

- The group would make proposal which would need to be reviewed and approved by the relevant EMA governance body before they are implemented

- **The set up of the group is in progress** – no specific timeline available yet as to when the group will become operational

- The scope of the Key User Group (KUG) will be to discuss RMS and OMS services, including OMS data management aspects e.g. Data Quality Standard monitoring and improvement

- OMS data stewards have already started to log data quality issues based on their experience with Change Request validations so far… *see snippet on the next slide*

Example data quality issues the group could discuss:

Data Issue #	Date of Issue	OMS data area	Issue title	Issue Description	Add - Country	For Decision	Standard to be applied
OMS-DI001	01/02/2018	AD address verification	Unwanted data enrichment	AD enriches address with District name which the user wants to remove.	DE	Option 1) We enrich the address as provided by AD Option 2) We Ignore AD and manually remove it every time Option 3) Reconfigure AD to ignore district names from the address reference data in all cases	
OMS-DI002	01/02/2018	AD address verification	Unwanted data enrichment	AD enriches address with State name which the user wants to remove.	DE	Option 1) We enrich the State as provided by AD Option 2) We Ignore AD and manually remove it every time Option 3) Reconfigure AD to ignore district names from the State reference data in all cases	
OMS-DI003	06/02/2018	AD address verification	FR characters in address	Postal service provider has removed some of the non EN characters from their addresses e.g. inverted commas on vowels, umlauts, etc. Informatica's address verification follows this approach. For more information click here.	FR	Option 1) We accept the data as verified by AD Option 2) We manually insert the non EN French characters	

- Data Quality standard is **used** by the EMA data stewards **to master organisation data** in OMS

- Users will find guidance useful when requesting additions and/or updates of organisations/locations in the OMS dictionary

- **OMS users are encouraged to become familiar with this guidance** (available on the OMS portal > documents tab)

- **Location details – main address**: submissions of MA applications with **enriched/transformed data** should not be a cause of rejection/invalidation of the application

- OMS Data Quality will improve over time

- Set up of OMS / RMS key user group is in progress

Thank you!
We will now take questions.

Follow us on 🐦 **@EMA_News**

Disclaimer

These PowerPoint slides are copyright of the European Medicines Agency. Reproduction is permitted provided the source is acknowledged.