How to best leverage your eDMS for the IDMP

An Ennov White Paper

Keywords

IDMP
eDMS, document-management
master-data-management, process optimization

Table of contents

1 Purpose of this document

One of the biggest challenges brought by IDMP into industry is the integration and standardization of data from multiples sources through new governance policies. Master Data Management (MDM) is largely considered to be a relevant approach to provide global framework and tools for this purpose.

In this context, electronic document management systems (eDMS) is a particularly significant source since most of the information required by IDMP exists in regulated documents, but is not readily usable.

This paper takes into consideration the ***interactions*** between eDMS and the IDMP, and provides insights and ***good practices*** to deal with data governance and IDMP compliance, addressing the following questions:

- How to manage the bi-directional flow of data to and from the IDMP reference data?
- How to transform your eDMS in order to best fit the IDMP needs?
- How to take into account the time factor considering current IDMP specification incompleteness and the progressive iterations of implementation in EU?
- What are the ideal interactions between eDMS, eSubmission and RIM?

2 What is IDMP?

2.1 IDMP in a glance

IDMP – Identification of Medicinal Products – is a new global ISO standard that defines ***data elements***, formats and ***terminology*** to uniquely identify and exchange information on medicine.

The primary goal is to simplify the exchange of information and enhance interoperability of systems between all stakeholders: health authorities (EMA, FDA or National Competent Authorities) and the industry.

It consists of five ISO IDMP standards which define data elements and structures for unique identification and exchange of regulated information on:

- substances (ISO 11238),
- pharmaceutical dose forms, units of presentation, routes of administration and packaging (ISO 11239),
- units of measurement (ISO 11240),
- regulated pharmaceutical product information (ISO 11616),
- regulated medicinal product information (ISO 11615).

For the pharmaceutical industry, being able to retrieve and manage such data, submit it to the authorities and maintain it over time constitutes as a significant challenge, which scope is still unknown to a large extent. This issue becomes more serious when considering that a significant part of the required information currently remains only in unstructured documents such as PDF or Office files.

2.2 IDMP governance

Dealing with IDMP compliance offers a great opportunity for process harmonization, the ins and outs are summarized in the following diagram.

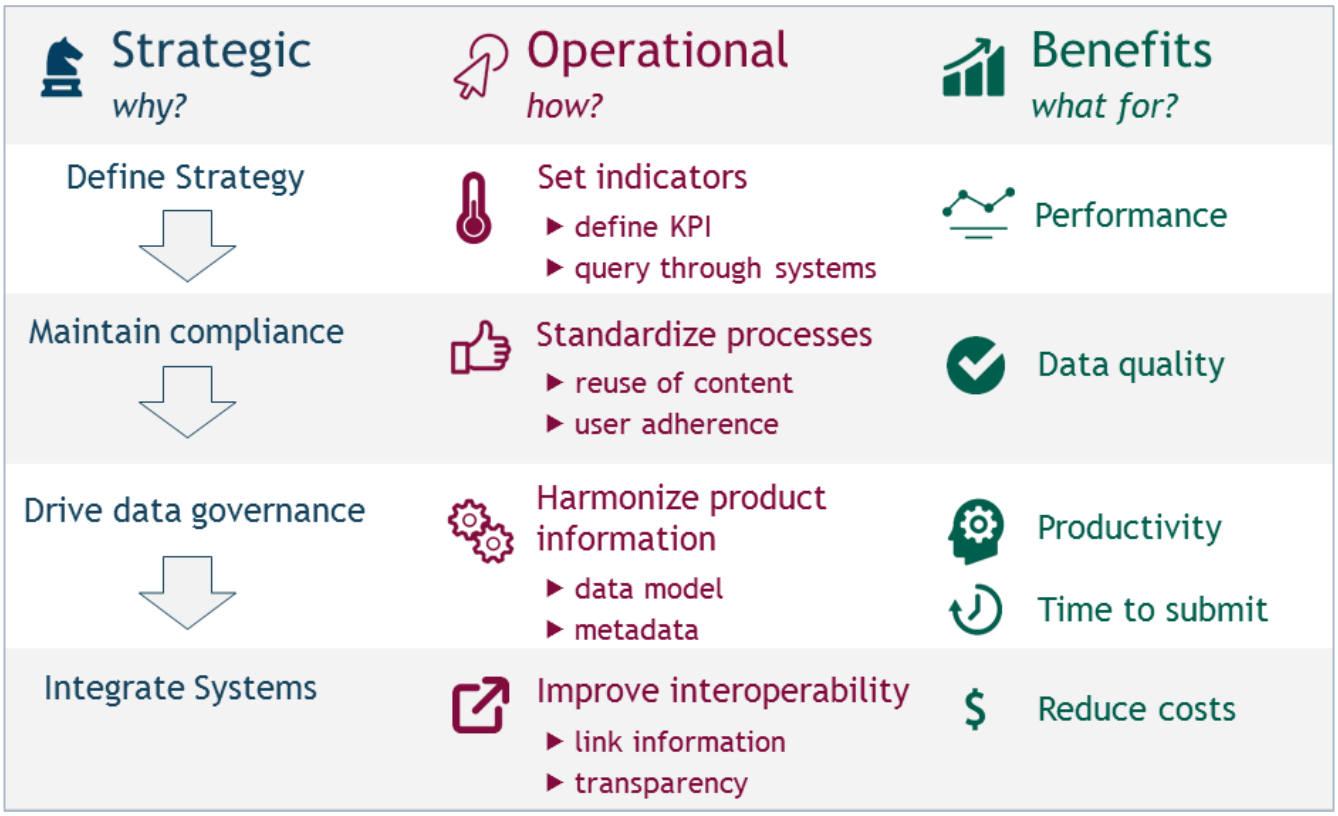


Fig. 1 IDMP governance overview

3 Documents in IDMP

3.1 eDMS – IDMP interactions

Electronic Document Management Systems (eDMS) serves as a major source of information in life science companies, and are likely to become a critical part of a global IDMP data feeding system, at least in the years to come. Addressing the IDMP would typically involve Master Data Management (MDM) in order to integrate data from multiple sources by means of a global data governance process.

In such a process, eDMS can both serves as an *input* and *output* of data.

As an ***input***: to feed the IDMP from both structured and unstructured data found within the documents:

- Structured: metadata associated to the document can help the integration process
- Unstructured: the textual content of the document itself, for example a significant amount of information can be found in the Summary of Product Characteristics (SmPC) but is often difficult to extract

As an ***output***: even correctly managed in an eDMS, documents would strongly benefit from being classified with IDMP IDs, and standard controlled terms instead of arbitrary code lists

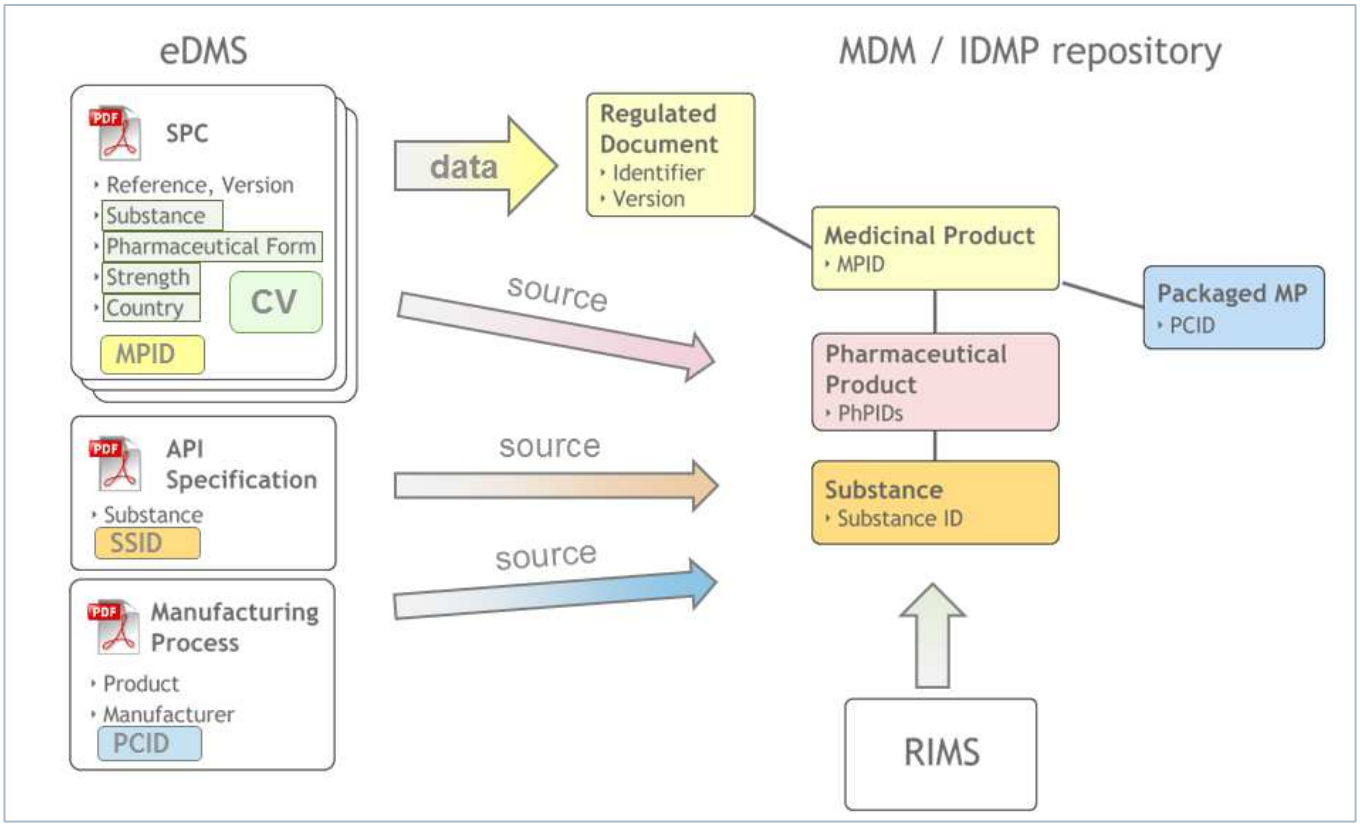


Fig. 2 Document / IDMP interaction

IDMP IDs are meant to identify various levels of information of a medicinal product:

- MPID: Medicinal Product ID identifies a unique product authorized in a country
- PhPID: Pharmaceutical Product IDs are a set of identifiers for identification for specific incomplete levels of information of a pharmaceutical product, for example when not knowing the strength of an active substance
- SSID: substance identifier
- PCID: packaged product identifier

Documents in the eDMS that are applicable to the specific final authorized product should be indexed with the corresponding MPID; documents that are specific to a substance should be indexed with the SSID or documents specific for packaging, indexed with the PCID.

But using IDMP ***controlled vocabularies*** in the eDMS have even greater benefits:

- to be able to link more easily any piece of IDMP information to a proper document source
- to allow reporting and analytics throughout the various systems that participate in defining and implementing global or local strategies, at the finest granularity of the documents and their versions.

3.2 IDMP "Regulated document"

The IDMP data model contains entity and data elements for "Regulated document" associated with the Medicinal Product. This would primarily be the SmPC which is submitted and approved by health authorities during the product registration process.

Each new version of this document is associated with a new version for the whole Medicinal Product data elements.

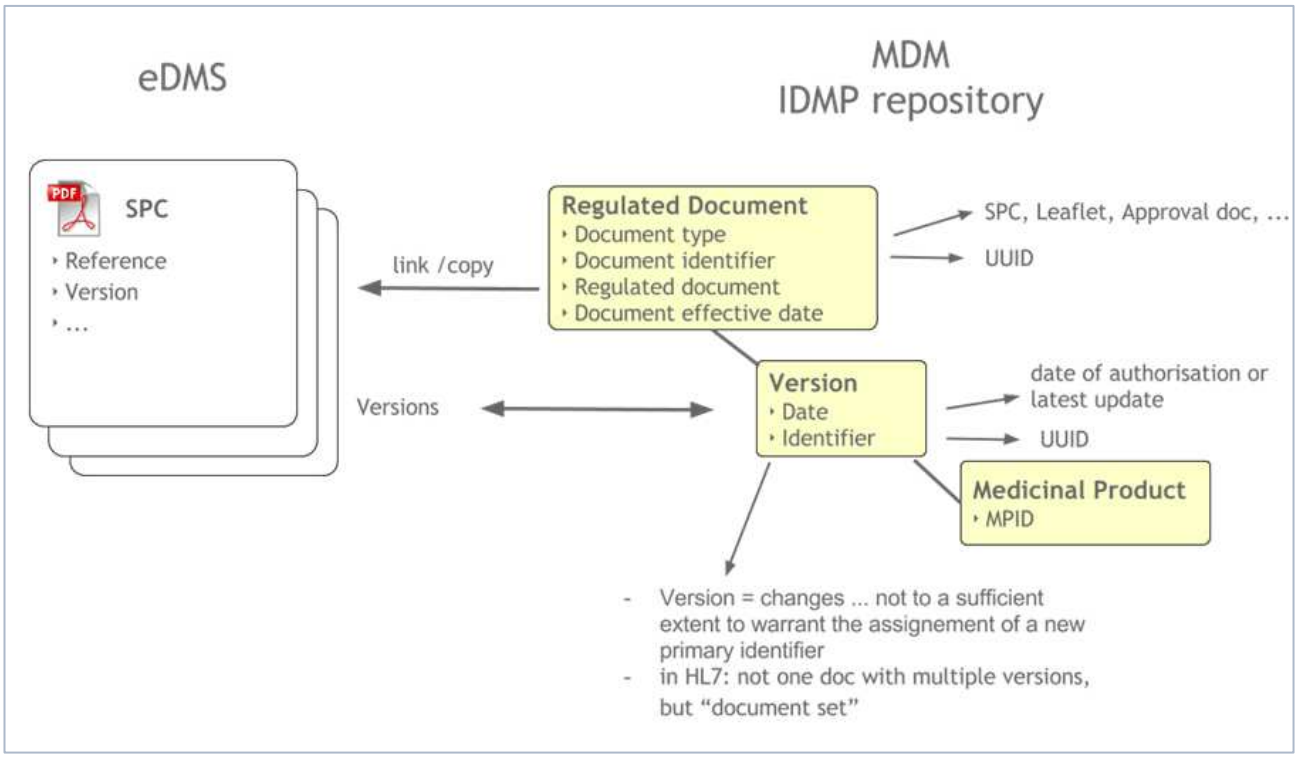


Fig. 3 IDMP Regulated document

Initially including such a regulated document in the IDMP repository doesn't constitute as a challenge within itself however, since these documents are usually already stored in one or more other systems, managing the changes and maintaining the up-to-date version can become problematic.

For example the SmPC is typically found in the following systems:

- eDMS, where the document authoring and life cycle is managed
- Publishing system, where the document is typically added to an eCTD application
- RIM, where the submission is tracked and has an approval status
- now IDMP, yet another copy of the document!

Multiple issues may result when simple duplication of the document file is made each time without having a proper governance schema, the most serious being not to be able to provide an accurate status for a given copy, reflecting both internal and external validation/approval.

The general answer to these issues relies on some core principles: responsibility, autonomy, linking:

- each system should have a clear purpose that does not overlap with the others
- it should be autonomous and manage only the metadata fields that are specific to its operations

- avoid copies/duplications of files and instead use ***document links***, which can be synchronized when a new version comes out. In some cases, an additional parameter for the link (current / fix) will maintain an up-to-date document (current) or a document from a fixed specific version (fix).

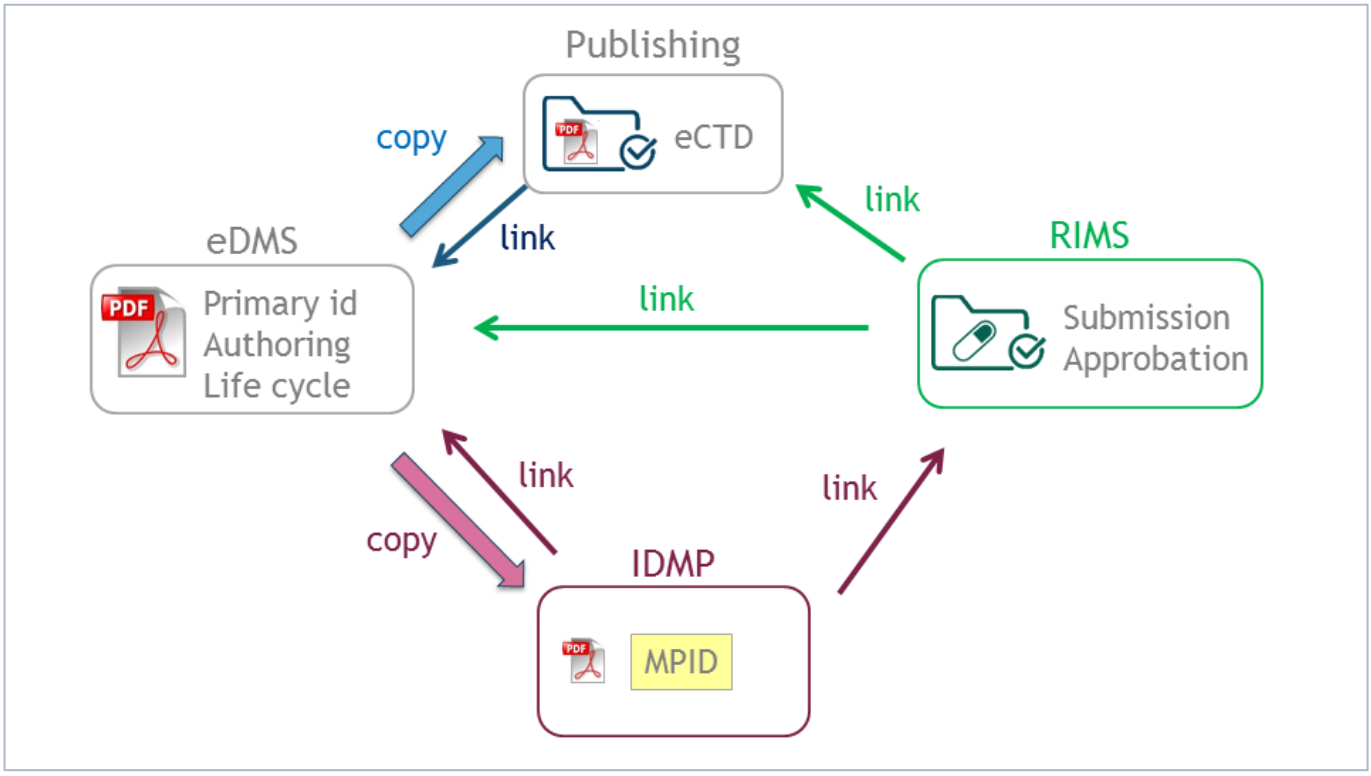


Fig. 4 Handle documents across multiple systems

3.3 Sourcing documents used in Text-Mining

Since a large part of IDMP data can be found in unstructured documents, and sometimes unfortunately only in those documents, an approach consists in extracting information from these sources. Such text mining methods would analyze the content of PDF or Office files and use patterns or statistical rules to extract and categorize specific data elements as defined in the IDMP.

In many cases, when considering the amount of related documents (several thousand or more), using such an approach appears to be the only realistic way to feed the IDMP data repository in a timely manner.

However, these techniques are also known to work with very variable rates of error which can significantly undermine the quality of the extracted data. Their efficiency is often measured by two indicators: accuracy and coverage. ***Accuracy*** shows how much of the extracted data elements are correct, or assigned to the correct entity. ***Coverage*** shows how much existing information in the documents was extracted.

As no information extraction system has 100% accuracy and 100% coverage (usually a compromise is made between the two), it is important to keep track of the ***source document*** for each extracted data

element, in order to allow for future verifications of this information, or to investigate on the extraction process mistakes.

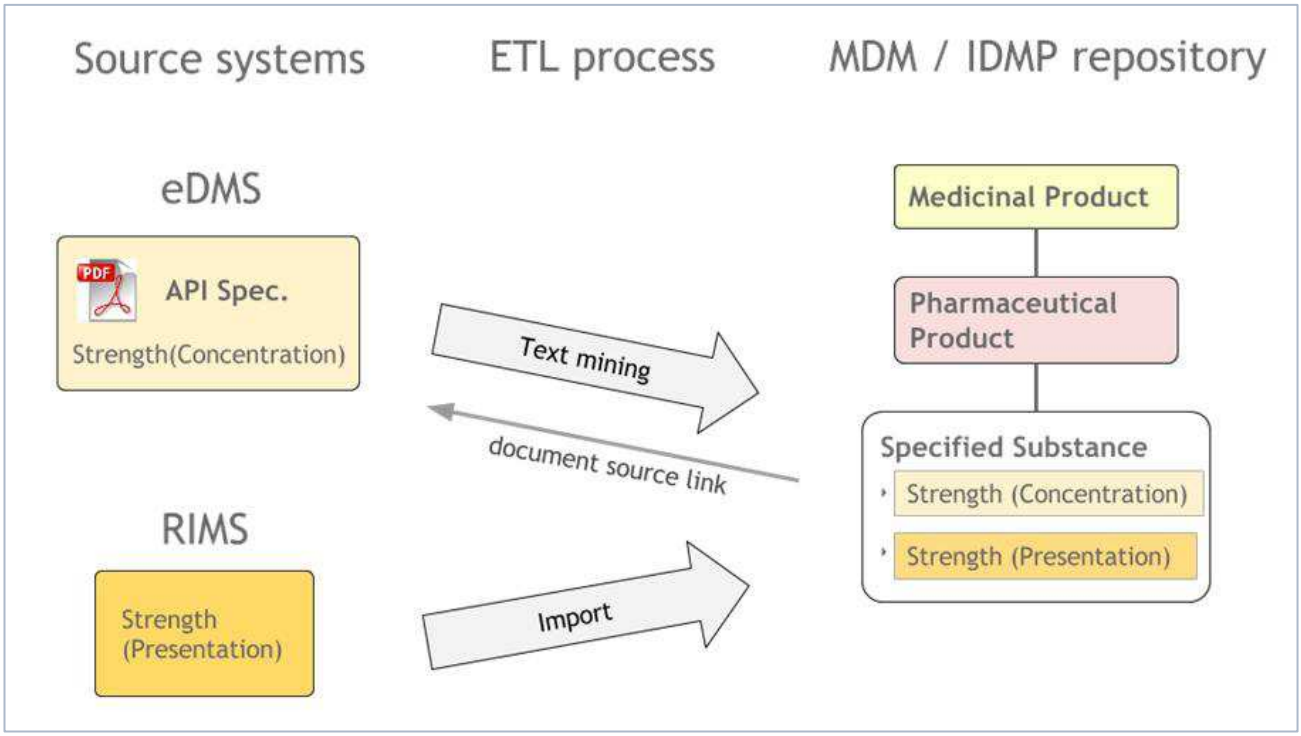


Fig. 5 Sourcing documents used in Text Mining

4 Integrating your eDMS with IDMP

In order to properly link the eDMS to the IDMP/MDM data repository, it seems inevitable to consider re-engineering of the eDMS to various extents.

4.1 Metadata oriented eDMS

Well-formed eDMS are metadata oriented where 'metadata' refers both to content metadata and context metadata.

Content metadata (*what*): they describe what the document is about, its theme, subject or any data element that the document specifically relates to, such as Product, Country, Substance, Manufacturer, etc. This is typically metadata relevant to the IDMP.

Context metadata (*who, where, when*): they describe the context of production and use of the document. It includes the document life cycle (authoring, review, approval, application date). It can also include *how* or *where* the document is used, but one should not go too far into this usage because it should not be the responsibility of the eDMS to track all possible usage of a document. For example tracking for which product and to which authority a document was submitted, and with which response status is more relevant to store in a RIM (Regulatory Information Management) system than in the eDMS.

Fig. 6 Document form in an eDMS

Managing metadata is essential for performing all essential operations in the eDMS:

- Identification of documents (using a reference and version)
- Browsing, listing: using document category and metadata columns
- Searching using any metadata criteria
- Reporting

The importance of metadata highlights the need to harmonize with the IDMP.

4.2 Challenges and opportunities for eDMS

The main challenges posed here, which turn out to be also a great opportunity for improvement, are described below.

1. metadata harmonization using IDMP Controlled Vocabularies
2. keep the eDMS focused on its primary goal which is to manage document content and life cycle, but not how the document is used in multiple other systems
3. be able to have in a document form in the eDMS, a link to the related IDMP entities, which then provide an extensive context of use for the document

Several options are available to operate the re-engineering of the document database, required by the above items 1 and 3:

- Full re-engineering
- Metadata re-engineering
- Document type re-engineering

4.3 Full eDMS re-engineering

The most exhaustive way to re-engineer the document database consists of extracting all data including document structure definition, attachments and metadata, modify structure definition and metadata in local files (for example excel), and import all this data to a new database.

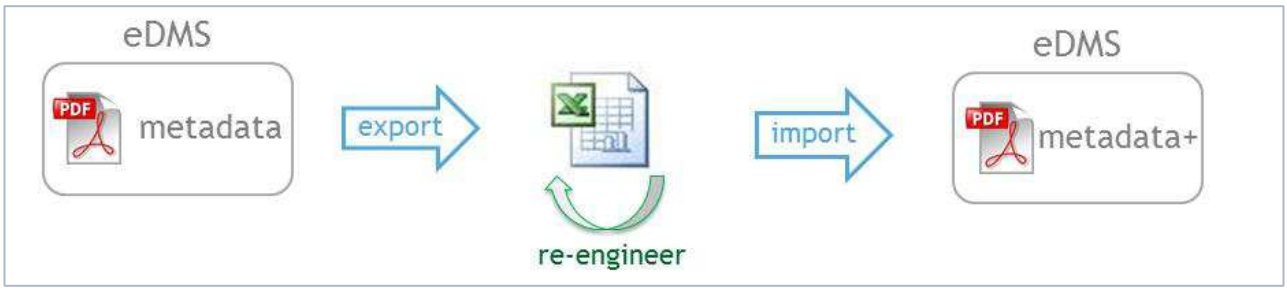


Fig. 7 Full database re-engineering

This strategy allows a complete restructuration of metadata and attachments since all data can be modified in excel before being imported. In particular, one may want to:

- clean data: remove duplicates or deprecated documents,
- reorganize classification for improved efficiency,
- replace uncontrolled terms by terms from standard controlled vocabularies provided by the IDMP
- add new metadata fields when needed, for example IDMP IDs or delete deprecated ones.

The export/import of all data can be a cumbersome process, depending on the size of the database. Most part of work load consists of metadata mapping and validating import data or procedure.

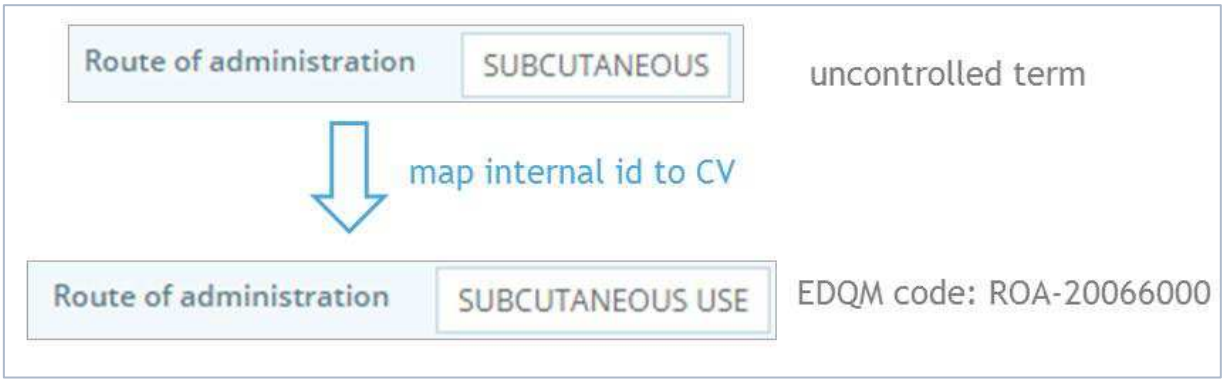
When changing metadata, be careful to keep original document main identifiers in order to preserve continuity and not disturb main user practices.

Reduce database lock time: when planning the operation on the live environment, consider the time needed to extract and re-engineer data, during which the live database may need to be locked.

4.4 eDMS metadata re-engineering

An alternative strategy to the previous one, involving a much lighter transition process would be to focus the changes to metadata contents only, keeping intact the main document structure.

This would allow for introduction of IDMP Controlled Vocabularies when a simple mapping is possible to the terms already used in metadata. Besides it would be difficult to manage situations where a used term is not precise enough and could thus be mapped to multiple controlled terms.



⊕ This method is much lighter since it doesn't imply actual change to documents but only metadata.

⊖ The scope of changes is limited to simple mappings and doesn't change the data model.

Example: re-engineering of a tree structure for products

In this use case, the need was to reorganize a molecule / product tree in order to insert a new level. The initial tree structure had the following levels: molecule, product/country, while the intended tree structure should have an additional intermediate level called 'Product', between the two.

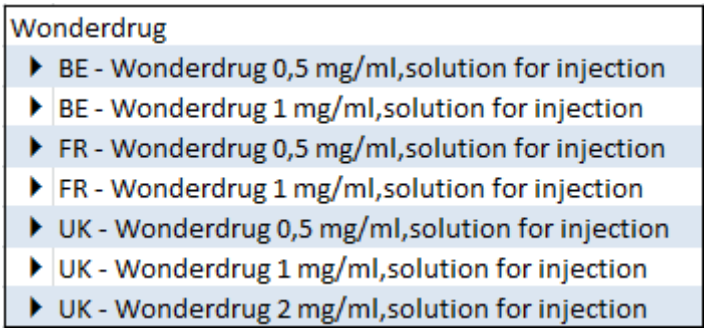


Fig. 8 Initial tree structure

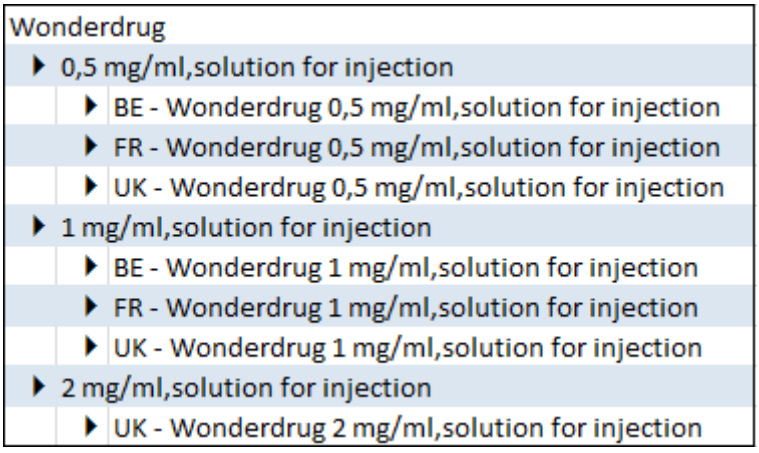


Fig. 9 Final tree structure

4.5 Document type re-engineering

This strategy can be used in addition to the light metadata re-engineering without being as complex as the full re-engineering. It consists of changing the document type definition in order to add new data field containing or linking to IDMP IDs.

This kind of evolution of document structure is usually a standard eDMS operation for the new documents being created, as shown in the diagram bellow, but it may not be for an application to all existing documents.

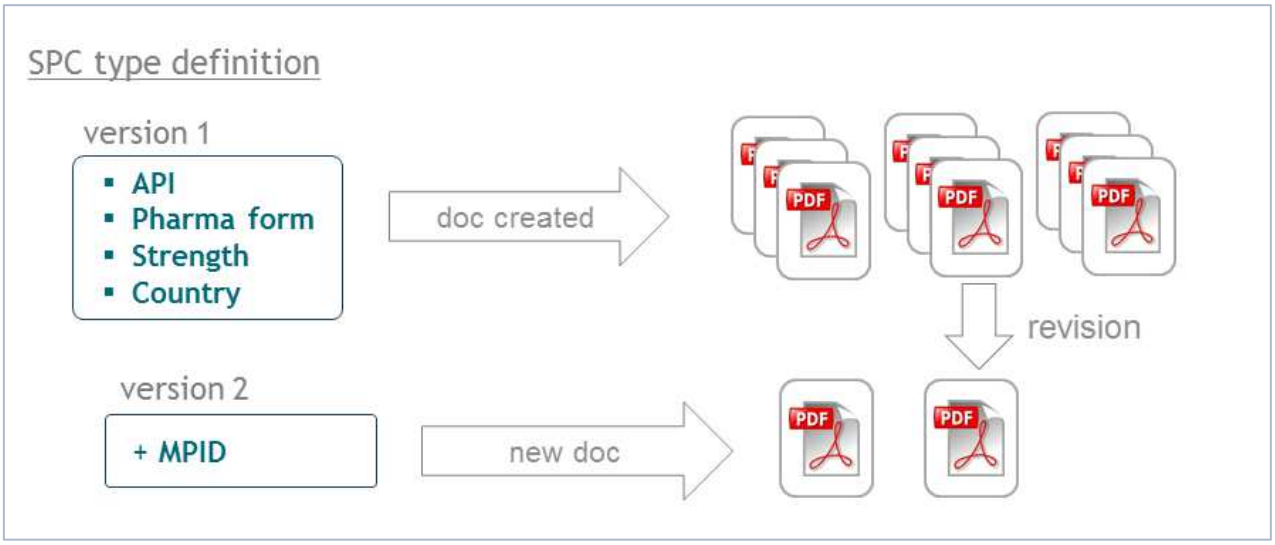


Fig. 10 Standard document type versioning

However the need here is to apply those changes to all existing documents, which can be performed in two steps:

1. apply type definition changes to all existing documents
2. assign new metadata values to documents (import or batch operation)

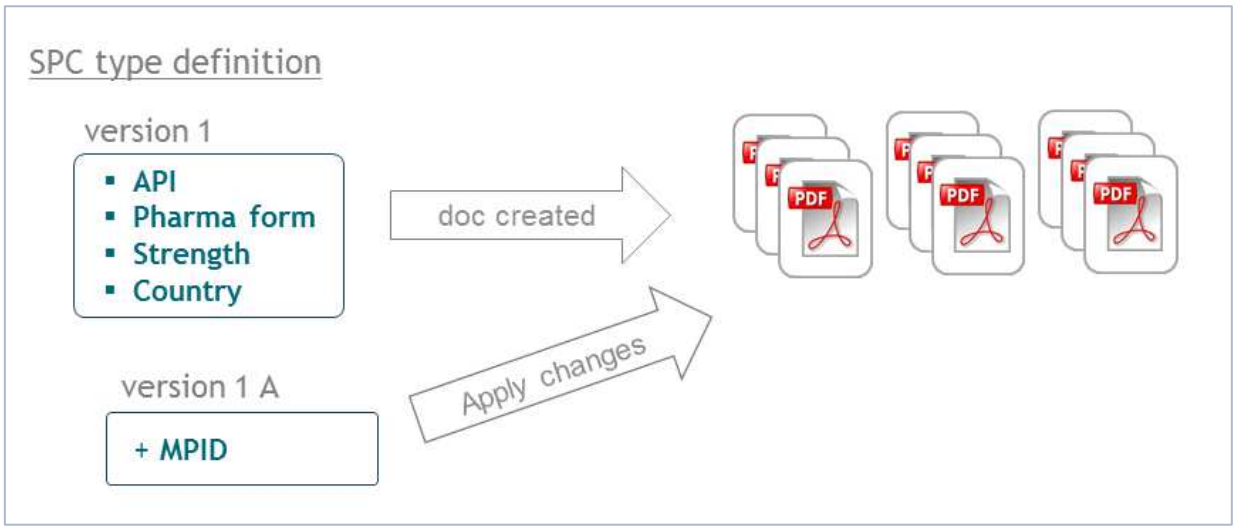


Fig. 11 Extended document type re-engineering

⊕ This method is a good complement to simple metadata re-engineering without being too complex. It addresses the most important issues of indexing documents with IDMP IDs and providing links.

⊖ Being able to modify document type definition to all existing document may not be an available functionality in your eDMS.

5 Conclusion

IDMP implementation is a global and complex process that will come through different iterations. One should carefully consider the evolution of IDMP requirements across time and regions (EU iterations, FDA) upon defining an internal strategy. At the same time, IDMP offers a unique opportunity to harmonize process and data in organizations.

Starting as soon as possible (now if possible!) to define strategies for re-engineering of your critical systems is strongly advised in order to leverage the changes brought by IDMP, in particular regarding cross linking and metadata.