

16 December 2021
EMA/447502/2021

European Medicines Regulatory Network Data Standardisation Strategy



Adoption by Big Data Steering Committee	16 September 2021
Adoption by European Network Data Board	8 October 2021
Endorsed by Heads of Medicines Agencies	24 November 2021
Endorsed by EMA Management Board	15-16 December 2021

See websites for contact details

Table of Contents

1. Introduction	3
2. Approach taken to develop the Strategy.....	3
3. Principles for data standardisation	4
4. Recommendations	5

1. Introduction

The amount of information submitted to regulatory authorities is creating an increasingly complex landscape for regulatory decision making. Current regulatory processes require the use of a diverse set of documents and data submission formats that have been developed independently over time. This requires substantial manual and technical efforts before the data can be further utilised for creating unique insights. This is recognised at an international level and various projects of standardisation are being undertaken through bodies such as the International Council for Harmonisation (ICH) and standard development organisations. It should be noted that standardisation work affecting the future landscape of healthcare data does not always include European involvement.

Until now regulatory procedures have been mainly document submission based leading to the assessment of information contained in documents rather than assessing the underlying data that were used to create those documents. The regulatory processes are shifting to assessing data rather than documents and looking at the potential secondary uses of the data collected to drive better regulatory decisions and improve public and animal health.

To simplify the process, over a decade ago, the EMRN (European Medicines Regulatory Network, hereafter also referred to as 'the Network') started requiring certain submissions to be compliant with recognised data standards. Data standards¹ provide a set of rules or conformance criteria that set out how information (data) should be structured, defined, formatted, or exchanged. Therefore, data standards are of great value to establish an efficient, predictable and consistent format allowing IT systems, interfaces and tools to receive, process, and make available data for re-use.²

Enabling and supporting harmonisation of data will reduce the administrative efforts currently required to receive, process, make available and reuse scientific data throughout the EU as well as in the global life science industry. It is therefore very impactful for the EMRN to take a leading role in supporting the adoption and adaptation of existing data standards and in supporting the development of new data standards. As acknowledged in the recommendations of the HMA-EMA Joint Big Data Task Force (BDTF) and the workplan of the HMA-EMA Joint Big Data Steering Group (BDSG), standardisation is a critical element for realising the full potential of (big) data and driving regulatory decisions³. Developing a Data Standardisation Strategy (hereafter named 'Strategy') for EMRN and its stakeholders' needs should serve as a roadmap to improve the way data on medicinal products is dealt with within the EU and to support the development of globally applicable standards for the human and veterinary regulatory domains. The strategy document will be maintained by being reviewed annually and revised as necessary according to new requirements and priorities that will arise in the future e.g. following the finalisation of the EU Veterinary big data strategy.

2. Approach taken to develop the Strategy

This strategy has been created through a set of activities to collect information across stakeholders about specific data needs that could be supported through the use of data standards. This strategy was also aligned with the goals and objectives identified in the EMRN Strategy 2025². The strategy activities are summarised below.

¹ Data standards are documented agreements developed by standardisation organisations that set out the representation, format, definition, structuring, tagging, transmission, manipulation, use, and management of data.

² [European Medicines Regulatory Network strategy to 2025](#)

³ [Big Data Steering Group \(BDSG\): 2020 report](#)

As an initial step, consultations were conducted over a period of 3 months to understand current practices applied by staff members for the management of scientific data and to create an inventory of the IT systems utilised for the daily work. These consultations were all transcribed and translated into use cases, which were extracted, logged and analysed.

Wider consultations were then held through the publication of an online public survey to all stakeholder groups that was followed up with a workshop with these stakeholders in order to collect further feedback and discuss the survey's findings. The outputs of the consultations and workshop were then transcribed and translated into additional use cases and analysed.

An analysis was also conducted on a selection of Standards Development Organisations (SDOs) known to develop healthcare related standards. This was done in order to develop an understanding of what data standards are already available for use and what new standards are under development. The identified set of standards were then assessed for applicability to the use cases collected and the use cases that are missing applicable standards were noted.

Data standards are made available in different forms that have implications on how they can be used and implemented. Some standards are published as logical models which provide definitions, values, and associations between data elements, they can also include business rules for checking the validity and conformance of the data. However, they do not always include a data exchange (message) format that enables the data to be transferred from one system to another, therefore an additional standard would be required for the data that conforms to the logical model data standard to be exchanged and shared.

Following the analysis of use cases and data standards a set of actions and activities were created that fall into one of the three following categories:

1. **Adopt:** Data standards that are already available but require adoption by the EMRN.
2. **Adapt:** Requirements which need an adaption of an available data standard to included additional requirements in a new version of the standard before it can be fully adopted.
3. **Develop:** Requirements that need an entirely new standard to be created.

These actions and activities were then developed as recommendations that are grouped into specific domains. The full set of use cases and further details of the analysis performed can be seen in the Data Standardisation Strategy technical annex. It is noted that it takes time and effort to create new standards and that adopting existing standards will normally have quicker implementation timeframes.

3. Principles for data standardisation

This strategy sets out the principles used to guide data standardisation efforts and the adoption of data standards by the EMRN. The principles applied to create the Strategy are adapted from the EU Data Strategy's concept of data to be Findable, Accessible, Interoperable and Reusable (FAIR).⁴ The FAIR principles main objective is to enhance the value, and thereby the usefulness of data

The Strategy further sets out the following principles that apply to creating and maintaining the strategy:

⁴ [EU Data Strategy](#)

1. Ensure the use of high-quality data standards developed by accredited SDOs in accordance with voluntary and consensus-based processes. Unless not compatible with EU legislation or otherwise impractical, these standards can be used in place of (unique) local or proprietary standards.
2. Contribute to the wider EU data and digital strategy including the creating of a European Health Data Space⁵ and follow the general provisions and requirements related to data sharing, re-use and data privacy laid out in the draft EU Data Governance Act.⁶
3. Ensure that data standards support data sharing & exchange, data protection and data ethics by design
4. Leverage collaboration and be informed by stakeholder needs.
5. Serve the EMRN: align and contribute to the EMRN Strategy to 2025.²
6. Consolidate the requirements outlined in regulations and guidelines through aligning with existing regulatory or health information technology initiatives and mandates.
7. Ensure the effectiveness and broad utility of data standards through the adoption or adaptation of other standards currently in use, when feasible.
8. Requirements should be assessed against what has already be implemented in order to prevent the use or creation of overlapping and competing standards. The assessment should include both human and veterinary needs so that where possible one standard that covers both domains can be used.

4. Recommendations

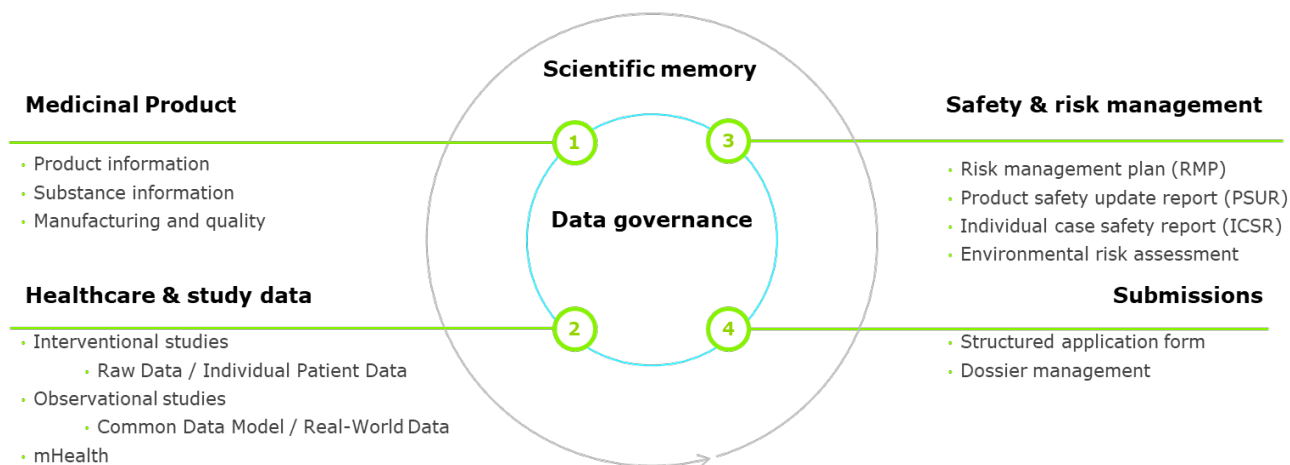
This strategy sets out four domains to organise the recommendations for the development and adoption of standards. The Medicinal Product, Healthcare & study data, Safety & risk management and Submissions domains were identified through the analysis work conducted to develop this strategy. The recommendations focus on where standardisation would bring the most benefits, fulfil the use cases captured and support the improvement of scientific memory in the EMRN. In addition to these domains the need for data governance has been identified to ensure that the strategy is implemented, regularly updated and that cross-domain work is coordinated.

Addressing knowledge management needs is relevant to the four main domains mentioned in this Strategy; products, healthcare data, safety & risk management and regulatory submissions, as well as other areas within the lifecycle of medicinal product development. Developing recommendations that would allow research to be captured and contextualised in such a way that it can be consistently interpreted is vital. Moreover, analysis of the different scientific data can be reviewed, and commonalities considered in relation to best practices and analytics so as to develop guidelines for a harmonised approach to scientific memory that respect the accompanying privacy and security considerations for personal health data.

⁵ [European Health Data Space](#)

⁶ [Data Governance Act](#)

Figure 1. Recommendation domains



1. **Medicinal Product.** The medicinal product domain covers product information, substance information and manufacturing and quality. Several projects are already actively working to create and implement ISO IDMP, using the HL7 FHIR standard for the exchange and publication of data concerning medicinal products for humans and veterinary use (e.g UPD, ePI & SPOR PMS).

Further work is needed to develop additional FHIR resources to structure more key product information that is not currently included in existing projects and ensure that this data is aligned with the ISO IDMP product and substance standards. This would enable greater availability of structured information which has a wide-reaching impact on patient health, the accuracy of product information and capability to mitigate potential supply chain shortages on an international and network level.

To support standardisation of substance information and manufacturing and quality, ISO IDMP specified substance group levels 2 and 4 could be implemented and adapted by the associated HL7 FHIR message exchange format. This will enable capturing of active substances and excipients in products in a structured format and will further support the inspections process and the ability to trace the origin of substances. In turn, the connection between product information and manufacturing information will facilitate faster responsiveness to potential safety, quality and supply chain issues.

Developing a standardised template for all inspections data, including good manufacturing practices (GMP), would make the identification of previous inspections and reports less time consuming. Moreover, providing the unique identifiers of manufacturing facilities will enable the creation of interactions between product quality and inspections of manufacturing sites (risk-based tracking). To assess product quality, an international standard for manufacturing quality raw data could be developed.

2. **Healthcare and study data.** This domain includes data from interventional studies, observational studies, the development of a common data model (CDM) for individual patient data and mHealth (mobile health) devices.

The ICH M11⁷ Expert working group is developing a standard that aims to structure the study protocol required for interventional studies. In addition to simplifying the submission process, the

⁷ [ICH M11: Clinical electronic Structured Harmonised Protocol \(CeSHarP\)](#)

structuring of this information will support the accessibility of the data and enable comparative analysis of trial data across different clinical trials. The work of ICH M11 should be further supported and further requirements should be considered for inclusion. In addition, the EMRN is currently investigating the benefits of receiving raw data underpinning interventional studies for which standards are already being piloted, including CDISC SDTM⁸ and ADaM⁹. These standards could be adopted for use once the results of these pilot projects are completed. The availability of the raw data is expected to support regulatory decision making through increased scrutiny, increased data quality and understanding of a marketing authorisation applications (MAAs) dossier.

Similar to clinical trials, observational study designs and results reporting could also be structured for better re-use and applicability of data across the Network. The current ongoing work related to creating a standard for clinical trial protocols, study design, conduct and study results could be extended to include observational studies.

In order to be able to use real world data (RWD) and metadata obtained from healthcare records and disease registries in a harmonised manner, a CDM standard needs to be developed and/or adopted. Stakeholders reported growing use of the OMOP CDM¹⁰ and its compatibility with FHIR. Use of a CDM will facilitate the access and reuse of data across countries.

With the increase in portable technology, mHealth represents a relatively new and growing medium of data collection. Having the ability to accurately utilise health data obtained from personal devices could support the enhancement of regulatory decision making by enriching data collected in clinical trials or healthcare record systems. This data would also provide valuable information in the derivation of safety and efficacy outcomes. However, this area requires further scoping to develop data standards that also respects the accompanying privacy and security considerations.

3. **Safety and risk management.** This domain covers Individual case safety reports (ICSR), product safety update reports (PSUR), environmental risk assessment and risk management plans (RMP).

Individual case safety report (ICSR) could take advantage of HL7 FHIR based messaging. A revision of the standard could also enable omics (genomics, proteomics and metabolomics) data of patients to be included in ICSRs.

The Network would benefit from an electronic product safety update report (PSUR) with structured information that follows the ICH E2C (R2) periodic benefit-risk evaluation report guidelines. This would facilitate tracking and linking product information and health information across data sets.

Currently environmental risk assessments are received as unstructured PDF files which limits the use and reuse of the data contained in these documents. In order to improve the usage of environmental risk assessment data the CDISC SDTM⁸ standard for data collection of risk assessment data would need to be reviewed and the adoption criteria be specified.

Risk management plans (RMPs) follow a paper template format and no electronic structured format currently exists. Creating electronic risk management plans (eRMPs) would enable linkage to registered studies and the related data that is stored in multiple systems. To achieve this, developing a new standard following the ICH E2E Pharmacovigilance Planning guideline and the EU Guideline on good pharmacovigilance practices (GVP) Module V – Risk management systems could be considered.

⁸ [CDISC Study Data Tabulation Model](#) (SDTM)

⁹ CDISC Analysis Data Model ([ADaM | CDISC](#))

¹⁰ [Observational Medical Outcomes Partnership \(OMOP\) CDM](#)

4. **Submissions** This domain covers the dossier management and the structured application form.


Compared to eCTD v3.2.2, eCTD v4.0 supports reuse of PDF documents across sequences and eCTD dossiers as well as allowing more flexibility for other data formats to be provided within the dossiers. The EMRN is in the process of transitioning towards the use of structured data which means that PDF files will become less common in the process of exchanging data. Therefore, other methods for data exchange such as HL7 FHIR messaging which would allow a more agile way of working could be considered to support timely exchange of data and documents should be reviewed.

The eAF currently uses a PDF format and will be replaced by a HL7 FHIR based standard for marketing authorisation applications (MAAs), renewals and variations to existing marketing authorisations. The eAF FHIR messages could also be used to cover the entire lifecycle of medicines and act as a means for populating regulatory databases (including human and veterinary data) with all the metadata required to run regulatory processes supported in all relevant regulatory systems.


Data governance. The mechanisms that are currently in place for the governance of data and submissions need to be reviewed in order to include the responsibilities for oversight of data standardisation work and for maintaining the EMRN strategy document going forward.

The Data Standardisation Strategy technical annex provides a representation of one way of prioritising the recommendations. However, this strategy does not include a roadmap for implementation and any prioritisation will need to include further assessment of the expected costs and benefits. This point underlines the critical importance of agreeing a clear data governance to take the strategy forward.


Recommendations for Medicinal Product, Healthcare and Study data




Product information: Plan further iterations of the electronic product information standard (ePI) to develop additional FHIR resources to support further structuring of information and ensure alignment with the ISO IDMP product and substance related standards.




Manufacturing and quality: To enable the assessments of product quality a group of experts should be tasked with developing a set of requirements for an international standard for raw quality data. In addition, further analysis is needed to determine how manufacturing, supply-chain traceability & inspections data can be standardised.



Observational studies: The standard being developed for clinical trial protocols and study design should be reviewed to see if can be extended to included observational studies. A CDM standard needs to be developed and/or adopted in order to facilitate use real world data (RWD) and metadata obtained from healthcare records and disease registries.




Interventional studies: A structured clinical trial protocol is being developed by ICH M11, this work should be supported by experts from the EU network to progress its development and include study design & reporting study results. Adopting relevant CDISC standards should be considered for collecting raw data.




mHealth: Further analysis is required in order to develop requirements for data standards that respect the accompanying privacy and security considerations. An expert group should perform this analysis.


Recommendations for Safety, risk management & Submissions




Risk management plan: A new data standard based on the published ICH E2E pharmacovigilance planning and good vigilance practice module V guidelines should be developed.




Product safety update report: An electronic product safety update report (PSUR) with structured information that follows the ICH E2C (R2) periodic benefit-risk evaluation should be developed.




Environmental risk assessment: The CDISC SDTM standard for data collection of risk assessment data should be reviewed and the adoption criteria be specified taking into account both the human and veterinary domains.



Individual case safety report: The individual case safety report (ICSR) standard could be revised to take advantage of HL7 FHIR based messaging and include patient omics data. Requirements for Omics data would need to be developed by an expert group before such a revision of the ICSR standard is undertaken.



Structured application form: The electronic application form FHIR messages currently being developed should be reviewed to see if they can be extended to support pre-application phase activities and include metadata to run regulatory processes.



Dossier management: The use of FHIR messaging for regulatory data and document exchange should be reviewed to see if it is the best option for the future.