European Federation of Pharmaceutical Industries and Associations

EFPIA Recommendations on a Connected Data System in Europe

Author: Digital Health WG **Date:** March 2022 **Version:** FINAL

HEADLINE MESSAGES

Objectives

- Connecting isolated data can unlock new insights in healthcare, leading to a healthier society. Enhanced data can help to design, develop and deliver better medicines that meet the needs of patients.
- Simple data connections include doctors within a country or between countries interacting to discuss diseases that are rarely seen. More complex approaches can lead to computer analysis of large volumes of data sets with the power to deliver novel insights on how to treat a disease or a patient.
- Setting aside pure research interests, interoperability will allow exchange of health records across borders and enable EU citizens to have more continuous, well-informed healthcare regardless of their point of interaction with healthcare system in EU Member States.
- A connected health data ecosystem has the potential to empower more effective and efficient research and development of new treatments and diagnostics. It would also ensure better planning and delivery of patient-centred care through personalised medicine. This, combined with value-based healthcare, can result in better allocation of resources and more sustainable healthcare systems.

Challenges

- Data can be contained in repositories, electronic records including clinical notes, electronic health records (EHR) and medical records (EMR), insurance claims, patient registries, records of patient reported outcomes/experiences, and continuous patient monitoring data (e.g. from apps and wearables). Unlocking the value of health data requires interoperability between different IT systems, providers, data sources and software.
- Conflicting national laws can be a barrier to data access and use. For example, the interpretation of the General Data Protection Regulation can vary between EU countries. Differing interpretations of data rights of citizens has the potential to undermine the key goal of the GDPR of harmonising rights and freedoms across Member States. EU Data Protection Supervisors must reach a common understanding of key GDPR terms if citizens are to enjoy the same rights across the EU.

Solutions

- It is essential to increase awareness and cooperation among all stakeholders and to develop a shared understanding of the relevant requirements in digital health.
- Connecting data requires an agreed format and approach (a common data model) to allow data to be accessed, pooled, compared and used, while ensuring privacy. EFPIA believes this is a critical step in enabling the benefits connected data can offer to the health of European citizens.

- Data does not always need to be transported or transformed to have broader use. It can remain distributed (federated) so that it can be used for its original purpose. Privacy by Design principles should apply.

Executive Summary

The value of data and of forming an inclusive European Health Data Space (EHDS)

The connection and flow of data across the European Union (EU) is a critical enabler of a healthcare system that values positive clinical and societal outcomes. A connected health data ecosystem has the potential to empower more effective and efficient research and development of new treatments and diagnostics. It would also ensure better planning and delivery of patient-centred care through personalised medicine.

Improved access to, and transmission of, health data could transform the pharmaceutical industry. Real world data can impact the entire drug development process, from discovery to validating medicines' real-world effectiveness and safety. This can catalyse the development of transformative therapies and the assessment of their effectiveness in broader populations over longer periods of time**.** For the research-based industry, access to data is critical at every step. From accelerating drug discovery to understanding patients' behaviours and the outcome of treatment, the availability of data is essential to testing hypotheses, identifying trends and assessing proposed treatment. Regulatory and legal uncertainty over the rules for pharmaceutical companies accessing, processing and sharing data would impact on the ability to innovate and respond to public health needs.

Interoperability to empower patients & support research, development and clinical care

Healthcare system information must be better connected. This will allow stakeholders to use this data for optimising and improving health outcomes. Interoperability is a critical enabler of the digital transformation of healthcare in Europe. **The EHDS should enable the exchange of data between different providers at the regional, national and pan-European level to support research, development and clinical care. This can be achieved by adopting interoperable quality and content standards, and by linking electronic health record systems and other sources of health data.**

Proposed solutions

As the EU considers the development of an EHDS, the concept of Federated Data Networks (FDN) may be particularly suited to the diverse needs of Member States with varying degrees of digital maturity and local regulations for data access. Essentially, an FDN is a managed architecture that allows for the sharing of mutual resources for Real World Data (RWD) use. This can unlock the power of data in primary or secondary care settings, in clinical care decision-making, and in research, whilst preserving the privacy of the RWD at a local level. In an FDN, data is not moved from its host source, though hybrid models can exist with local and central data hosting. The research question or query moves to where the data is originally hosted, with results aggregated centrally or delivered to the researcher, applying Privacy by Design principles.

The use of principles such as FAIR (findable, accessible, interoperable, and reusable) data, provides the framework for exploiting the benefits of an FDN. This is enabled by Common Data Models (CDMs), open Application Programming Interfaces (APIs), standardised communication protocols, metadata, standardised analytical tools and fit-for-purpose methodologies. A CDM is essentially a construct, a means to an end to help organise RWD into a common structure, formats, and terminologies across diverse, heterogeneous, and multiple source datasets. It addresses a central need to be able to curate

data for analysis on a contemporaneous and continuous basis (not on a per study basis) or for large-scale, geographically diverse, network studies of multiple data sources.

This paper elaborates on various existing CDMs, with reference to European Medicines Agency's own evaluation of CDMs from a regulatory perspective which includes guiding principles that can be utilised more broadly. **EFPIA acknowledges and advocates for a broader recognition by the European Commission of the benefits of CDMs and supporting an FDN, with the aim of reducing the overall latency and resource requirements for conducting research at scale and ensuring quality more rapidly.** EFPIA strives to contribute as a key stakeholder to the debate and decision on which CDM(s) should be agreed internationally. Our common goal is to improve the health of European citizens through the power of available data.

I. Maximising the Value of Data

Data, and digital tools such as Artificial Intelligence (AI), are essential components of the pharmaceutical strategy for Europe. The European Commission's 'Updating the 2020 New Industrial Strategy: Building a stronger Single Market for Europe's recovery' includes important initiatives on resilience, the green agenda and digitalisation. The EU pharmaceutical strategy should be aligned with the New Industrial Strategy. It must focus on ensuring that Europe remains an innovator and world leader in the development and manufacture of medicines and active ingredients, with the objective of ensuring fair access for patients. Tomorrow's breakthroughs in medicine rely on the appropriate regulatory framework that supports today's research needs.

EFPIA recognises that an innovation-friendly ecosystem will rely on a robust data infrastructure and governance model that would allow the pharmaceutical industry and academia to operate with a transparent framework for increased access and analysis of data for secondary use.

The connection and flow of data across the EU is a critical enabler of a healthcare system that values positive clinical and societal outcomes. A connected health data ecosystem has the potential to empower more effective and efficient research and development of new treatments and diagnostics. It would also ensure better planning and delivery of patient-centred care through personalised medicine. This, combined with value-based healthcare, can result in better allocation of resources and more sustainable healthcare systems. A successful EHDS can also inform policies that support improved overall health outcomes.

A significant lesson of the COVID-19 pandemic has been the need for real world, observational data, in the right hands, to identify insights from the past and present and inform the future of pandemic management. The pandemic has acted as a catalyst for numerous developments in the clinical care setting, responding to the need to provide care remotely. Telemedicine has been growing in use over time, as well as routine use in synchronous, asynchronous and/or remote patient monitoring. Lockdown and resource restrictions by stretched healthcare systems provided an opportunity for expanded adoption that should continue after the pandemic.

Improved flow of data and use of digital technologies enables patients to become more engaged in their care. Patients increased use of mobile technology and ability to aggregate and share data about their own health is moving the patient centre stage. It may promote a better quality of patient-clinician interactions and information exchange, providing valuable insights into patient perspectives that might otherwise not be reported.

One of the main benefits of measures facilitating the use of health data for healthcare is the ability to conduct primary and secondary research that could lead to new innovative, transformative therapies for patients. This can facilitate research into the genetic basis for disease and develop targeted therapies to address areas of high unmet need, for example. **Access to, and transmission of health data, could transform drug development through the use of RWD as a complement to clinical trial data, speeding development of potentially transformative therapies.** This is critical throughout the medicine's lifecycle:

- to design and run new trials through feasibility simulation and support for recruitment,
- for evidence generation across the medicines' lifecycle,
- to inform dynamic price-setting in relation to the value of medicines,

- to optimise appropriate use of medicines in daily practice, in epidemiology and medicine safety surveillance,
- to support formal assessment by regulatory and Health Technology Assessment (HTA) bodies, providing insights for life science research and effective medicines utilisation.

Analysis of data informs research and development strategies, and it can improve the potential value of products in development by identifying the diseases and subgroups most likely to benefit from a new treatment. **If the EHDS and the rules surrounding access to the data are not carefully thought through, with the involvement of all stakeholders, there could be unintended consequences that limit the utility of the data for developing innovative medicines.**

Wider application of digital health services could allow for a health system that is centred on people's individual needs and preferences, with important implications for how we measure health system performance. Addressing roadblocks in the patient journey and removing duplicate interventions can accelerate patient care, reduce unwarranted costs and potentially improve health outcomes. Integrated care has the potential to increase continuity, improve efficiency, support patients' empowerment, and foster health system sustainability and resilience, while reducing waiting times. For some diseases and conditions, especially the most complex ones, consistency of outcomes and the ability to compare them can help improve care coordination. Digital health services and Electronic Health Records are also crucial tools to strengthen care coordination and service integration, as well as to improve self-management through patient access to their own health data.

Collaborative EHDS

The EHDS should be created via a process of coordinated action with a participatory governance model involving all stakeholders. At all times, access to the data and accompanying restrictions should be proportionate and duly justified. The conditions for use and exchange of health data should be clear, not too prescriptive, fit for purpose and should take into consideration the level of access to the data as well as other safeguards.

For the research-based industry, access to data is critical at each step of medicine development. From accelerating drug discovery to understanding patients' behaviours, the availability and quality of data is vital to testing hypotheses, identifying trends and assessing proposed treatments. Regulatory and legal uncertainty over the rules for pharmaceutical companies to access, process and share the data would impact the ability to innovate and respond to public health needs.

For example, varying interpretations of the General Data Protection Regulation (GDPR)[1] from Data Protection Authorities within Member States present challenges for clinical development of innovative medicines. Conflicting interpretations of Article 9 of the GDPR, and the additional limitations on processing of health and genomic data that Member States have enacted under this article, cause significant delays in study start-up and patient enrolment. Some Member States take the position that the only lawful basis for processing health data is when individuals have given their consent for its collection and use. Other Member States take the position that processing this health data, when necessary for scientific research, is lawful. There has been no way to manage this conflict to date other than using multiple informed consent forms and long negotiations with multiple ethics committees.

[1] https://eur-lex.europa.eu/eli/reg/2016/679/oj

Conflicting interpretations of the definition of various entities under GDPR – specifically, who is a data controller and who is a processor – also cause major delays in study start-up. Pharmaceutical companies must contract with clinical trial sites, and, due to the differences in interpretation, some ask to be considered co-controllers of the study data. Pharmaceutical companies are then entangled in long, difficult contract negotiations. This delays study initiation at these sites and access to clinical trials for European patients in need, potentially delaying beneficial therapies from reaching the market. These delays are due to interpretation of European privacy law.

It is important to note that currently, there is no common European interpretation of what constitutes 'sufficient anonymisation,' or 'pseudonymisation, 'or 'secondary use' of data. It is important that all stakeholders are involved in the discussions of how these terms should be defined with respect to the EHDS, as the definition of sufficient anonymisation depends on who is sharing the data, who sees it, and what they are doing with the data. Efforts to define a one-size-fits-all approach to anonymisation can lead not to a lack of sufficient anonymisation, but to *over* anonymisation that renders the data not useful and, at worst, potentially inaccurate. This can inhibit research and development and potentially lead to inaccurate conclusions about the effect of a novel medicine in development. Additionally, there are significant costs to the research enterprise if data are stripped of information that is critical to assessing the risk/benefit or safety/efficacy of a potential product. Numerous data-sharing initiatives exist. If everyone does not apply the same standard, it makes it impossible for researchers to use the resulting data and draw conclusions from it. This is a major impediment to advancing the field of data sharing and research.

II. Interoperability to enable exchange of data

Connected EHDS

Healthcare system information must be better connected. Interoperability is a critical enabler of the digital transformation of healthcare in Europe. Setting aside pure research interests, interoperability will allow exchange of health records across borders and enable EU citizens to have more continuous, well-informed healthcare regardless of their point of interaction with healthcare system in EU Member States. Healthcare systems should also benefit from easier access to past medical history, thus reducing inefficiency.

In addition to these societal benefits, medical research will benefit enormously from the Commission's plans to implement exchange formats for Electronic Health Records (EHRs). EFPIA believes the strongest argument for adoption of interoperability standards within the EHDS is to facilitate greater access to cross-border health data in the region. This should be done using a common exchange format with visibility of technical specifications and access controls to ensure confidence in data safety, privacy and security. A common entry and access route will encourage innovation, lowering the barrier of entry and ensuring a level playing field for data owners, regardless of size or resources.

Lack of interconnectedness leads to a lack of access by patients and healthcare professionals. This prevents data being used for evidence-based decision making and for scientific research. Unlocking the value of health data requires interoperability between different IT systems, providers, data sources and software. This is essential to provide a holistic view of patients' health, enable personalised healthcare, boost the flow of research and facilitate the creation of learning healthcare systems. The future should see connected data systems where clinics securely share vital patient information to enhance individuals' care; clinical trial data from around the world is shared across borders to develop

new safe and effective medicines more quickly; and RWD is pooled to extract real-world evidence on how interventions add value.

The EHDS should enable the exchange of data between different providers at the regional, national, and pan-European level, promoting data exchange to support research, development, and clinical care. This can be achieved by adopting interoperable quality and content standards and linking electronic health record systems and other sources of health data. As the EU comprises 27 Member States, broad network research requires porous digital borders, as is the case for data portability to support patient mobility. This necessitates federated approaches to overcome the challenges, particularly in allowing remote, secure interrogation (but not movement) of data.

The governance framework should prioritise standardisation needs and improve data interoperability. Ideally, this will serve to enable seamless connection of data sources in the context of the EHDS, enabling high quality insights to be derived by data partners and will work to build overall trust in the data ecosystem.

III. Use cases showcasing the potential of databases and a CDM

Databases are vital in providing accurate, timely and comprehensive health care services to patients. Therefore, it is critical that the data is of high quality, well structured, as accurate as possible, and standardised to a Common Data Model (CDM) to facilitate transfer or pooling.

An example, showcasing the potential of existing databases, is the jointly-led initiative between academia and the pharmaceutical industry called EUMelaReg that aims to build a European wide infrastructure for the collection and integration of RWD from melanoma patients throughout Europe. The objective is to exploit existing registries in Europe. In countries where melanoma registries do not exist, the technical infrastructure to build these out will be provided. Currently, 13 countries are committed to the creation of a multinational registry with an aim to optimise the management of malignant melanoma in Europe.

Availability of a significant amount of high-quality, interconnected data about the individual is also crucial for the application of reliable and trustworthy AI in healthcare. The PULSE-AI project aims to calculate a patient's individualised risk of Atrial Fibrillation (AF) – which is the most common arrhythmia – to target which patients should be tested based on patient characteristics and temporal trends in patient data. For that purpose, an algorithm calculating the risk of AF was created. PULSE-AI is a double (baseline and time-varying) neural network that is a learnt model based on one million primary care patients in the UK. It is a validated innovative approach for the identification of patients at risk of AF and their subsequent diagnosis. The PULSE-AI tool identifies one in nine patients deemed to be high-risk with undiagnosed AF and is being embedded within a national electronic health record system.

So far, the system was applied in the UK where data in primary care is rich and deep allowing a highly predictive neural network to be developed. The AI algorithm was derived and validated in a large UK dataset with over three million patients. Although the approach is UK specific, risk is defined based on patient characteristics and as such has the potential to be applied to other localities, given the connectivity between primary and secondary care data is ensured. A lack of individualised patient data impacts the ability of the algorithm to directly define risk. In the UK, it has been shown that high quality

connected data on individualised patient's health will improve accuracy which could prevent 2,000 strokes and saves over £40 million across the health and social care economy.

IV. The concept of the Federated Data Network (FDN)

As the EU considers the development of an EHDS, the concept of a FDN has been raised by various stakeholders. Essentially, an FDN (sometimes referred to a distributed data network) is a managed architecture that allows for the sharing of mutual resources for Real World Data (RWD) use. This can be used in primary or secondary care settings and in clinical care decision-making, as well as in research, whilst preserving the primacy of the RWD at a local level. Data is not moved from host source, (though hybrid models can exist with local and central data hosting). The research question or query moves to where the data is originally hosted, with results aggregated centrally or delivered to the researcher. It is a socio-technical construct, including the technical architecture and tools to facilitate the network, with governance aspects based on agreements, codes of conduct and adherence to legal and privacy requirements (such as GDPR). Privacy by Design facilitates the community's use of the data in the network.

The technical architecture in a FDN allows for source data to remain secure behind its sociotechnical firewalls, i.e. technical security through to approvals and ethical oversight. Web-based tools and technologies mean source data can be analysed where it is, especially if it is organised in a way that facilitates this – for example, via a CDM, supported by central portals and management, inclusive of metadata-driven catalogues.

Though different in sociotechnical aspects, FDNs such as the FDA's SENTINEL, PCORNET, OHDSI (Observational Health Data Science and Informatics), ADVANCE, ConcePTION, TriNetX, EHDEN (European Health Data & Evidence Network) and DARWIN EU (Data Analysis and Real World Interrogation Network) and the EHDS already exist or are being built in open science or commercial communities. Use of such principles as FAIR (findable, accessible, interoperable, and reusable) data, provide the framework for exploiting the benefits of an FDN, enabled by the use of CDMs, metadata, standardised analytical tools and fit-for-purpose methodologies[2][3]

[2] Go-FAIR; FAIR Principles; https://www.go-fair.org/fair-principles/; accessed 7th March 2021
[3] Wilkinson MD, et al; The FAIR Guiding Principles for scientific data management and stewardship; *Scientific Data* 2016; 3: 160018

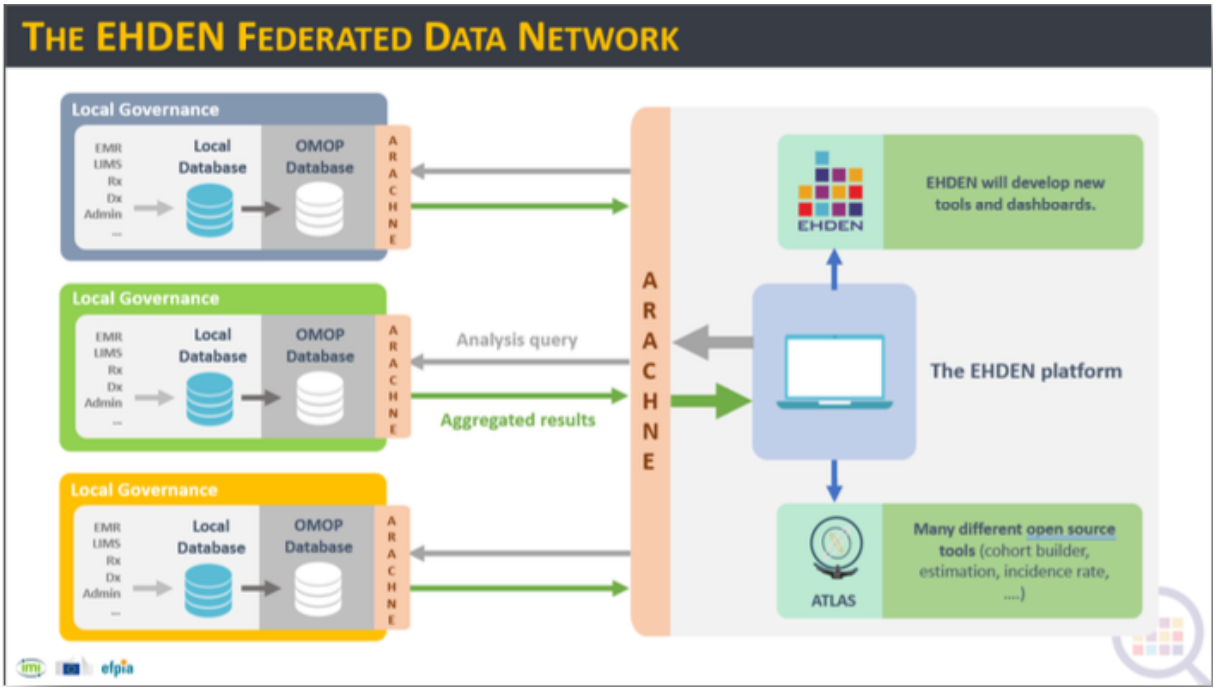

Figure 1: FDN framework explained through EHDEN example

The FDN framework may particularly suit the EU's need across diverse Member States with varying degrees of digital maturity. Ultimately, a hybrid of centralised and federated approaches is likely. There may be technical and methodological reasons for using a centralised data hosting architecture, albeit within a federated network, such as central databases or data lakes. For the needs of European healthcare systems, clinical care and research, a mixed ecology of architectures will most probably support a diversity of needs and use cases. While centralised data architectures have existed for some time (whether databases or data lakes) this has been prohibitive in expense and resources, especially at scale. Centralised architectures come with increasing scrutiny and legal, governance and privacy restrictions, and are more complicated for the data custodian or controller, and for researchers. Certainly, within the European landscape, increasing responsibilities cause additional overhead for central architectures. Moreover, the need for transparency in the use of real-world health data means open science socio-technical architectures are required, rather than proprietary and/or black box approaches.

This may be related to privacy concerns, but also reflects the need of regulatory authorities to understand the analytical path from source data to evidence. Moreover, being able to utilise a CDM to harmonise languages is an advantage in network, multi-site studies across borders. Some have expressed concern regarding the contemporaneous nature of the data being mapped, i.e. how often is it refreshed following the original mapping to a CDM. This is highly dependent on the source data custodian's refresh cycle, and this can vary between 24 hours and weeks or months. However, many aspects of the mapping refresh, including for iterations of the CDM itself, can and are being increasingly automated.

Access to data is more about the terms of access, rather than direct access to RWD. For example, the administrative burden for approvals and contracts in conducting real world (especially network studies) is significant. Though clear governance requirements are a necessity, there need to be mechanisms to address the administrative burden associated with them. Models such as Data Permit Authorities (DPAs) (e.g. FinData or the French Health Data Hub) may point to a potential solution.

V. Standardisation for efficient RWD analysis

What is a Common Data Model – and why use one?

A CDM is essentially a construct, a means to an end to help organise RWD into a common structure, formats, and terminologies across diverse, heterogeneous, and multiple source datasets. It addresses a central need to be able to curate data for analysis on a contemporaneous and continuous basis (not on a per study basis) or for large-scale, geographically diverse, network studies of multiple data sources.[4] This reduces the overall latency and resource requirements for conducting research at scale. It ensures quality more rapidly than other methods, especially in supporting an FDN (though CDMs can be used for centralised databases too). The mapping process itself inherently incorporates data quality audit of both the source and the CDM-mapped data, with iterative stages per mapping cycle and over time.

A key concept is the need to standardise data which has been collected, stored, and curated differently, whether in an institution, or across data sources, to an international scale. The Clinical Data Interchange Standards Consortium (CDISC) standard, utilised especially for Randomised Clinical Trials is a CDM, enabling regulatory authorities such as the FDA to receive, analyse and opine on diverse studies across the pharmaceutical industry. The SENTINEL CDM was designed to address the need to do the same for RWD with an emphasis on regulatory pharmacovigilance in the United States. The Observational Health Data Sciences and Informatics (OHDSI) Observational Medical Outcomes Partnership (OMOP) CDM is facilitating a global open science network.

Standardisation can ensure that diverse data is broadly mapped to common schema, ontologies, and vocabularies, for instance with OMOP, SNOMED[5]. Furthermore, it can support the use of standardised analytical methods and tools, on top of the CDM mapped data, following extraction, transformation, and loading (ETL) into the CDM. Exemplars of studies, such as drug utilisation, safety, regulatory, and HTA studies, lend themselves to greater consistency and commonality of methodological approach through standardised analytics on top of a CDM (such as SENTINEL ARIA's system). The use of a CDM can underpin the operation of an FDN via facilitation of distributed data querying across multiple data sources, all mapped to the same CDM, from studies through to federated predictive analytics.

[4] OHDSI; HL7 International and OHDSI announce collaboration to provide single common data model for sharing information in clinical care and observational research; https://www.ohdsi.org/ohdsi-hl7-collaboration/; accessed 7th March 2021

[5] SNOMED International determines global standards for health terms, an essential part of improving the health of humankind.

Reviews and comparisons of differing CDMs exist, but the EMA's own evaluation of CDMs from a regulatory perspective features guiding principles that can be utilised more broadly[6]:

Structure:

- A CDM:
  - can be defined as a mechanism by which the raw data are standardised to a common structure, format, and terminology independently from any particular study in order to allow a combined analysis across several databases/datasets.
  - should not be considered independently of its ecosystem which incorporates standardised applications, tools and methods and a governance structure.
  - requires that the ability to access source data be retained.
  - should be the simplest that achieves security, validity, and data sufficiency.
  - should be intuitive and easy to understand.
  - should enable rapid answers to urgent questions when required, be efficient and feasible.

Operation/Governance:

- The CDM:
  - governance model must respect data privacy obligations across all data partners and regions.
  - should be built with sustainability as a priority.
  - development should maximally utilise data partners' expertise. The CDM must be agreed by and accepted by the participating data partners.
  - must have version control.
  - should be dynamic, extendable and learn from experience.
  - should have no/low barriers to entry for new data to be used in an FDN.
  - Should have a value package that is clear to data partners.

Quality of Evidence Generation:

- The CDM:
  - must operationalise reliability, validity and data integrity by building clear and consistent business rules around transformation of data across multiple databases. Where divergence is unavoidable this should be recorded.
  - focus should be on data characterisation to understand if the data is fit for purpose.
  - should be transparent on how data is defined, how it is measured, and should incorporate and document its corresponding validation.
  - should allow transparency and reproducibility of data, tools and study design to facilitate credible and robust evidence across multiple datasets.

Utility:

- The CDM:
  - should provide a common set of baseline concepts which should enable flexibility when required and meet the needs of potential users.
  - should map all the concepts that are commonly used in safety and effectiveness studies to the CDM to maximise regulatory and value assessment utility.
  - should address recognised use cases for which an established need is present.

[6] Dodd C, Gini R, et al; D7.5 Report on existing common data models and proposals for ConcePTION; https://bit.ly/3H2SkXH, accessed 9th November 2021

Examples of Common Data Models

Currently only two CDMs cover the majority of these principal requirements at significant scale: SENTINEL's in the United States, and the OMOP CDM internationally. For Europe, there is little utilisation of the SENTINEL CDM but expanding adoption of the OMOP CDM.

Via the IMI EHDEN project, 12 EFPIA companies are providing funding of EUR15 million through 'in-kind' contributions (alongside EU funding of EUR14 million) over the duration of the project to accelerate utilisation of the OMOP CDM across the European region, with more than 20 IMI projects utilising this CDM.

In recent years the OHDSI OMOP CDM has become an international standard for working with RWD in RWE generation, with more than 2 billion health records mapped to the OMOP CDM globally, and a growing body of literature from international studies, all characterised by their scale and speed, while preserving quality. As well as running SENTINEL, the FDA is also funding OHDSI through the Center for Biologics Evaluation and Research (CBER) for biologics and vaccines pharmacovigilance, and the DARWIN EU will include the OMOP CDM framework.

The open science approach within OHDSI was demonstrated during the COVID-19 pandemic, through a study-a-thon and continuing research protocols, as was the TriNetX commercial FDN, through its international research studies. Such approaches responded to the need for the right data to be in the right place at the right time to address crucial questions during a public health emergency. In contrast, more traditional approaches would likely still have not reported results, especially for large scale studies with multiple data sources across the European region.

Comparisons of common data models exist, as discussed in the EMA report on CDMs in 2018[7]:

Table 1: Comparison of three CDMs.

[7] European Medicines Agency; A Common Data Model for Europe? – Why? Which? How?; London 2018

FDA SENTINEL	PCORnet	OMOP
Focused use (pharmacovigilance)	Clinical care emphasis	Broad use cases
US based	US based	Global use
Distributed data network with data queries run locally	Predominately EHR data	Broad, comprehensive model to incorporate claims data, EHRs and surveys
Predominately US claims data, minimum, but expanding EHR data	Principle of minimal mapping	Substantial mapping of content and concepts to standardise multiple different coding systems
Strict version control	Strict version control	Strict version control
Built upon principle of minimal mapping and no derived values	Flexibility for individual data partner to add data/ domains to local CDMs	Iterative development by community for data/ domains additions in global CDM
Source data retained		Source data retained
Extendable	Based on SENTINEL CDM	Extendable

The OMOP CDM was designed from the ground up for research purposes, initially in North America and with an emphasis on epidemiology, utilising e.g. US claims data. It has been expanded over the following years, both in terms of data types incorporated and study types supported, as well as for geographies. More recently, this has included regulatory use cases and, currently, developments to enable HTA studies or precision medicine use cases. Due to the open science emphasis of OHDSI, there is a focus on transparency, replication of results, and development of methodologies for fit-for-purpose RWE generation and observational research.

The OMOP CDM and OHDSI framework do not support every conceivable use case, and likely a mixed ecology of applications, methods and tools will be required to do so. This reflects the reality of working in the real-world setting. Further interoperability, e.g. between HL7 FHIR (for facilitating health data exchange) and OMOP CDM (designed for RWD analysis), to support outcomes research, is being addressed and accelerated with the recent announcement of a global collaboration.

Skilled and knowledgeable epidemiologists, statisticians or data analysts with multi-year experience of the OMOP CDM, mapping datasets and analysis using the OHDSI framework are a prerequisite now for some positions. A helpful example of this is a company's ability to make quicker decisions about the feasibility of being able to conduct a substantive study in collaboration with regulatory authorities. This work would be assisted by a transparent, reproducible methodology federation. The use of the OMOP CDM is now also supporting initiatives focused on therapeutic areas within the company as it proceeds to expand its collaboration with potential data partners.

In 2020, UCB presented at the OHDSI symposium on their work with the SME, The Hyve, in the Netherlands, on adaptation of the ATLAS tool to incorporate a new cohort sampling function and an extended patient profile view. This reinforces the extensible nature of powerful tools operating on top

of the OMOP CDM, and this work has been further incorporated into the OHDSI community for global utilisation[8].

Other projects in IMI have developed CDMs, such as ADVANCE and ConcePTION, in vaccines and pregnancy research, respectively, with the former using a CSV format CDM and Jerboa data processing software and R scripts, and the latter using a syntactic model.[9]

Conclusion

EFPIA believes that the European Commission needs to agree and endorse a common data model approach in order to unlock the value of otherwise isolated data sets. There are many organisations working on the best approach. EFPIA remains available to work alongside others as a key stakeholder in determining which CDM(s) should be embraced internationally to foster a harmonised approach. Our common goal is to improve the health of European citizens through the power of available data.[10]

[8] UCB & The Hyve; New ATLAS features: Cohort Sample and Profile timelines; https://forums.ohdsi.org/t/new-atlas-features-cohort-sample-and-profile-timeline/12135; accessed 8th March 2021

[9] Sturkenboom M, Braeye T, et al; ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccination; Vaccine 2020; 38(2): B8-B21

[10] EFPIA position on EHDS