20 December 2021
EMA/109720/2020
Data Analytics and Methods Task Force

EU Big Data Stakeholder Forum 2021

Report of the EU Big Data Stakeholder Forum - 7 December 2021

Table of contents

Introduction

The pace of change in technology, in data science and in healthcare is unprecedented and brings new opportunities to better the health of European citizens through innovation, better regulation, and learning from healthcare data. The European Commission's Digital and Data Strategies, adopted in early 2020, set out a bold vision for health data including the creation of a European Health Data Space. In 2022, this will be further realised through a legislative proposal for the European Health Data Space. The European Commission's Pharmaceutical strategy for Europe is also being developed with legislative proposed anticipated.

In this context, the second workplan of the HMA-EMA joint Big Data Steering Group (BDSG) 2021-2023 was published in August 2021.

To foster collaboration and to engage stakeholders, the European Medicines Regulatory Network organised the second annual Big Data Multi-Stakeholder Form. The Forum informed stakeholders on the delivery of the data pillar of the Network Strategy 2025 via the HMA-EMA joint BDSG workplan, provided an opportunity to listen to stakeholders' views and feedback and discussed areas for collaboration.

The Forum was opened and introduced by Emer Cooke (EMA Executive Director), Karl Broisch (HMA Management Group Chairperson and President of BfArM), Andrzej Rys (Director - Health Systems, Medical Products and Innovation, DG SANTE, EC) and Marco Greco (President, EPF).

One hundred and sixty-five registers participants attended the Forum by webinar and additionally the Forum was streamed live on the web.

This report offers a high-level summary of the Forum discussions.

Session 1: Report on implementation of the HMA-EMA Big Data recommendations

The Forum started with an overview of the Big Data Steering Group 2021 deliverables and plans for 2022 presented by Dr Peter Arlett (EMA).

By delivering the vision of a regulatory system able to integrate Big Data into its assessment and decision making, we can better support the development of innovated medicines, deliver life-saving treatments to patients more quickly and optimise the safe and effective use of medicines through measurement of a products performance on the market.

The second BDSG workplan was published in August 2021 to continue to progress the activities launched in 2020 and to address new topics. Significant progress continued in 2021 to enable the data transformation of the EU regulatory network and the delivery of these activities remains on track.

To enable the use and establish the value of Big Data, EMA is following key implementation principles:

- Step wise approach and use-case driven

- Learn from experience & stakeholders

- Demonstrate value

- Align with EU & international initiatives

- Future proofed - agile and evolve over time

Key achievements and stakeholder engagements of 2021 and a preview of 2022 were presented before being discussed in more details in section 2 of the forum.

Progress on the veterinary Big Data was presented by Ilaria Del Seppia (EMA). In 2021, a first Veterinary Big Data Stakeholder Forum was held. Despite current efforts are focused on Regulation EU 2019/6 (VMP-Reg) implementation, Big Data is already being used in the veterinary domain. Aspects to be addressed include building a data mind-set, establishing an international cooperation forum, determining data governance, funding and resources. An EU Veterinary Big Data strategy proposal will be developed across the Veterinary Regulatory Network and a draft proposal on EU Veterinary Big Data Strategy was shared with HMA in November 2021. For 2022, the EU Veterinary Big Data Strategy is planned for adoption, a training curriculum will be defined and thematic workshops with NCAs and stakeholders will be organised.

> Top priorities for 2022 indicated by the big data forum participants:
>
> **1. Data quality and representativeness, 2. DARWIN EU and 3. Regulatory processes**

Session 2: Big Data - enabling use and establishing value

In the second session, for BDSG recommendations, the achievements, future highlights, benefits and opportunities for stakeholder engagement were presented by BDSG members followed by a discussion with selected panellists' representatives of key stakeholders' groups.

DARWIN EU ®

DARWIN EU ® has seen significant progress in 2021, moving from a vision to launching the tender to select the coordination centre that will establish and operate the DARWIN EU ® network on behalf of the EU regulatory network. Including key stakeholder representatives, the DARWIN EU ® Advisory Board has also been formed, and provides strategic advice and recommendations to the project team on establishing DARWIN EU ® and its use of the EHDS, ensured coordination and alignment with relevant European and EU Member State initiatives and policies and supporting communication on DARWIN EU ®.

In early 2022, the DARWIN EU ® coordination centre will be appointed, and progress will enable the EU regulatory network to have more routine access to RWE (together with training, processes, catalogues of studies and data sources) and start conducting studies for decision making. DARWIN EU ® will also support the European Health Data Space (EHDS)2 pilot and engagement with stakeholders will be intensified.

DARWIN EU ® is supported by stakeholders to improve the decision-making process by EMA and NCA committees by generating high-quality real-world evidence. Stakeholders also see the potential to support decisions beyond regulations e.g. HTA bodies and payers.

Key recommendations from panellists and stakeholders:

- Patients highlighted the need to **focus beyond electronic heath record (HER) and consider other type of data** to be included in DARWIN EU ®. DARWIN EU ® will provide important **value for rare diseases**.

- Health Care Professionals called for **Patient Reported Outcomes (PROs) to be included in the scope of DARWIN EU ®** to capture data such as on quality of life (that are not captured in EHRs) and the need to ensure interoperability.

- Data permit authorities recognized that DARWIN EU ® will facilitate access to more data sources, enabling interoperability of system across Europe. **Valuing and rewarding the work done by data partners, HCP and health care systems to collect data will help ensure quality from the start** of the RWE generation process.

- While HTA and EU regulators have complementary responsibilities, the need for clinical evidence on use, and benefits and risks of medicines are common, and **synergies and benefits for HTA can be seen** for DARWIN EU ®.

- Industry recognized the health and societal benefits of DARWIN EU ® for regulatory decision making, and the **need for transparency and collaboration** (on analysis processes, methods, results).

Data quality & representativeness

In late 2021, EMA initiated a project to deliver the first version of the EU data quality framework for data used for regulatory purpose. Collaboration with 'TEHDAS (Towards a European Health Data Space) has also been established and will intensify next year. Finally, EMA initiated plans to discuss with a focus group the qualification process through Scientific Advice Working Party.

In 2022, informed by a stakeholder workshop, a first version of an EU data quality framework will be launched and recommendations to strengthen the process to advise on data qualification will be proposed.

Stakeholders recognized the importance of this recommendations and how it will establish norms for data quality including for real world data, provide clarity for data collectors and data managers on quality requirements, support the assessment of data sources and of study results and increase the impact of individual data sources through their qualification (for regulatory decision-making).

Key recommendations from panellists and stakeholders:

- Patient monitoring and patient empowerment can drive better health outcomes. **High-quality and representative data is relevant to patients and will enable their engagement**.

- **HCP expertise can be leveraged** to navigate this complex and rapidly evolving data environment.

- TEHDAS highlighted the need to improve data quality, the collection of metadata and of patient data to address fragmented access to healthcare services. **Timely provision and collection of data is critical and should be considered.**

- For contract research organisations (CROs), data definitions, reproducibility, and archiving **need clarity on regulatory requirements**. For large clinical trials across multiple countries or regions, convergence of requirements is needed.

- Academia considers that **strengthening the data qualification advice process will increase trust, transparency, and interoperability** of big data systems. A regulatory framework for data quality is need as well as guidelines (e.g. data quality assessment). Data quality indicators should be included in the RWD public catalogue.

- Industry underlined that **data quality will drive up trust in real world data and Big Data** and will ensure comparability. **Certification of data source should be considered,** and **data governance will be key**.

Data discoverability

Informed by consultation with stakeholders (including a dedicated workshop, surveys, and presentations) the first list of real-world metadata for regulatory purpose was created in 2021 and included the definitions of key concept, the first list of criteria for inclusion of databases in the real-world metadata catalogue (i.e. a data source catalogue) and the relevant associated processes. Together with the above, a proof-of-concept for the RW metadata catalogue was implemented and will serve as the basis for the delivery of the definitive EU regulatory real-world metadata catalogue.

In 2022, the list of real-world metadata for regulatory purpose (v.1.0) will be agreed and published for stakeholders, together with a good practice guide. The real-world metadata catalogue will also be made available to stakeholders and will start to be populated with the agreed real-world metadata. Finally, the enhanced catalogue for observational studies will be launched.

Collaboration with stakeholders will continue and alignment with other initiatives will intensify, including with TEHDAS.

There was general support for this work that will add additional benefits to stakeholders increasing knowledge on data, transparency, reproducibility, access and use of data and ultimately supporting the ability to judge the evidentiary value of observational studies and real-world data sources.

Key recommendations from panellists and stakeholders:

- Patients highlighted the need to **consider the sensitivity of data** (e.g. genetics data) and the opportunity to collaborate with patient organisations.

- HCPs called for **adequate processes and tools to locate, access and share data**, for **accountability to maintain data over time** and for a **transparent risk assessment process on the data source** (quality + representativity). **Discoverability should be a priority considered by design, including access.**

- ENCePP also recognised the need to **keep catalogues up to date** (completeness and timeliness) and for continued collaboration to advance the ongoing work.

- Academia noted the **need for more research on research methodologies** (e.g. research for causality still needed) and that progress in data discoverability will **support acceptability of RWE** and **demonstrate their evidentiary value.**

- Industry underlined that it will help to integrate learning from similar studies and the need for sustainable processes.

Skills

Following the adoption of the Biostatistics and Pharmacoepidemiology curricula in 2020, in 2021 an additional Data Science curriculum was adopted by the Big Data Steering Group. A Survey of EU Network skills was also delivered to inform on future needs and prioritisation.

In 2022, procurement for external training providers to deliver these 3 training curricula will be launched, ultimately rolling out the trainings to the EU Network via the EU Network Training Centre. Consultation on possible additional training needs (e.g. pharmacogenomics) will be organized.

Stakeholders recognised the foundational aspect of this work to increase the level of expertise on pharmacoepidemiological methods and biostatistics and to support EU regulatory decision making processes from the definition of research questions and elaboration of study protocols to the interpretation of study results and strengths and weaknesses of using certain types of data sources and research methodologies.

Key recommendations from panellists and stakeholders:

- Patients consider **development of skills part of establishing trust** in what we do and that training should **go beyond regulators**.

- Academia recommended to **focus on the tasks need for studies rather than general training for staff development** (e.g. data curation, study design, understanding of data sources, best data source or methods to use) and to identify ways to **attract specialists and new talents** in the drug development domain.

- HCP called for **consideration for the role of General Practitioners** (GPs) playing a key role in collecting data but for which data collection is not part of current culture hence training would be beneficial to understand the bigger picture. Opportunity exists such as collaboration with WONCA.

- Regulators recognised the **value across the EU regulatory decision-making process** of investing in training and skills to **increase consistency and quality of** regulatory advice and assessment. **Regulatory training should be rolled out**.

Regulatory processes for data

In 2021, PRAC established processes for RWD access to routinely support decision-making, based on earlier piloting. Several RWD use cases have been confirmed and proof of concepts and pilots for delivery of RWE generated by the network started with the Scientific Advice Working Party (SAWP), the Paediatric Committee (PDCO), the Orphan Committee (COMP) and the Committee on Advanced Therapies (CAT).

Critical to the development of processes, to guidance for industry, and to delivery of data-driven decisions is to systematically learn from applications to the Network that include Big Data. In November 2021, a 'learnings initiative' workshop was held and included the results of the CHMP review of Real World Evidence (RWE) in Marketing Authorisation Applications (MAA) and extensions of indications (Flynn et al., 2021) published in 2021 and past piloting of RWD analysis in committee decision-making (notably with the PRAC). This workshop will further inform process and guidance improvement in 2022-2023.

In 2002, qualitative aspects of the RWE submitted in applications will be investigated to characterize their contribution to the benefit risk evaluation and decision making, more pilots will be started to

deliver RWE to CHMP, national competent authorities, HTA bodies and payers and pharmacogenomic use cases will be discussed with the BDSG.

Activities in this area are one of the top priorities for stakeholders to better understand the use and characteristics of RWE in applications submitted to regulators by medicines developers and support individual product submission and guidelines by learning from current experience, to ultimately enable more use of RWD and accelerate the availability of treatments for patients.

Key recommendations from panellists and stakeholders:

- Patients called for **PROs,** including **quality of life data to be considered** as benefits will go beyond regulators and for **more collaboration with rare disease patient organisations and ERNs**.

- The regulatory network wanted to expand into **hospital data** and more data **representativeness of the EU population** (notably more data from Southern and Eastern Europe).

- HCPs highlighted the need to **consider data on non-prescribed medicines** and the potential to **leverage data from community pharmacies across Europe** for potential use cases (including treatment adherence/non-adherence).

- Industry stands ready for collaboration and consultation and called for **more guidelines and best practices while also balancing flexibility and regulatory certainty**, preferably in line with international progress.

- SMEs called for **learning from the pilots to be discussed with Stakeholders**.

Regulatory capability to analyse data

In 2021, the lessons learned from the clinical trial raw data pre-pilot have been finalised and the Advisory Group on Raw Data has been established to inform the full pilot of analysis of clinical trials raw data in MAAs that will be executed next year. To complement the work on raw data, a pilot has also been initiated to leverage unstructured information in documents submitted to regulators as part of scientific advice requests and MAAs. This will enable easier and more accurate identification of patterns across procedures to enable research and support generation of guidance.

The BDSG is exploring how data analysis clusters of excellence at national level can be fostered, including through mutual support and sharing of good practice (in the area of data access, legal aspects, capabilities, infrastructure, methods development and Artificial Intelligence - AI) with the drafting of a discussion paper on Clusters of Excellence (planned to be published in 2022).

Following the AI workshop in 2021, work to draft guidance on AI in medicines regulation will be initiated in 2022.

In 2022 stakeholders will also be engaged in a workshop on clinical trials raw data, including the Network's review on interim lessons learned from the raw data pilots of MAAs. Finally, the business case and practicalities for analysis of raw data from non-clinical studies will be explored.

Key recommendations from panellists and stakeholders:

- Patients called for **results of the pilots to be discussed with stakeholders** and for collaboration with patients on the guidance on AI.

- HCP highlighted the need to **comply with data protection and for more work on PROs**.

- Regulator remarked that AI has multiple touchpoints with regulation and **guidelines are needed considering the rapid evolution of this field**.

- Industry supported the need for guidance on AI, including the interface with new EU legislation and how deal with algorithms that are not yet regulated. **AI is an excellent area for multi-stakeholder collaboration**.

Delivery of expert advice

In 2021, the EMA Management Board agreed with the creation of a Methodology working party to modernise and strengthen the delivery of expert advice and methodologies including on Big Data. Methodology guidelines have continued to be rolled-out starting with a substantial revision of the ENCePP guide on pharmacoepidemiological methods published in July 2021 and the new CHMP Guideline on Registry based studies.

In 2022, a roadmap for the development of comprehensive guidance across data and methods will be developed and the Methodology working party will be established, including a broad range of expertise (i.e. biostatistics, real world evidence, advanced analytics, PK/PD, extrapolation, modelling and simulation, GCP and omics). Further research on methods will take place including via the call from Horizon Europe Cluster 1 Health on 'New methods for the effective use of RWD in regulatory decision-making'.

These are additional foundational activities that will further contribute to the conduct and assessment of high-quality studies to generate evidence for regulatory decision making.

Key recommendations from panellists and stakeholders:

- Patients highlighted the need to **consider involving patients in the design of registries**. More work is also required to **better explain to patients why their data is needed and how their data is used**. **A trusted third-party model is a possible solution to facilitate sharing of data**.

- HCPs noted the need for m**ore education and training, not only for the EU regulatory network but also for a wider stakeholder community**. There is also an **important link between data and communication**, and **contextualising data use in relevant health examples** like vaccines has been very helpful during the pandemic.

- Academia supported the creation of the methodology working party with interdisciplinary skills.

- Industry is willing to share experience on RWE for decision making and noted that **qualification of EHR sources** will be an important aspect to consider for the future.

Data governance

To guide stakeholders, to support compliance and to enable public health research, a question and answer document on data protection in the context of secondary use of healthcare data has been prepared and will likely be published in 2022, subject to the publication of the anticipated guidance from the European Data Protection Board. Together with this, EMA will also publish data protection training via the EU Network Training Centre to support the EU Regulatory Network.

In addition, BDSG held exploratory discussions on data ethics and prioritised to engage with ongoing initiatives (e.g. TEHDAS).

In 2022, the BDSG will continue to prepare for the future EHDS and a technical workshop with TEHDAS on data governance is proposed to ensure alignment, collaboration, and preparedness.

This work will enable consistent application of EU data protection law in the context of the secondary use of health data and ensure we operate within an ethical framework for big data research considering ongoing initiatives and expert advice. The importance of robust data governance is recognised by all stakeholders. Cross-fertilization between EU projects and initiatives and international fora will help consistency of application of data protection law and help address uncertainties.

Key recommendations from panellists and stakeholders:

- Consumers highlighted the **need for harmonisation in data protection legislation implementation** and the need to consider the **role of ethics committees and ethical approaches to conduct research with big data** e.g. different role of informed consent with secondary use of data.

- Establishment of data permit authorities provides an **opportunity to further support harmonisation** (e.g. data access, data availability, data quality, metadata) and interoperability. **Sustainable funding to support these activities is needed**.

- TEHDAS called for **governance to be overarching, harmonised and robust to support trust**. Collaboration with TEHDAS will inform the work of the BDSG.

- A data protection representative noted the uncertainties among stakeholders with secondary use of data (e.g. different implementation, data transfers needs, anonymisation). GDPR is generic and not specific for health data. **We need a practical toolkit on data protection and to bring data protection authorities to multi-stakeholder platforms.**

International initiatives

Following extensive stakeholder consultation in 2021, the data standardisation strategy for medicines regulation (including Big Data) has been agreed and [published] .

International collaboration on RWE has continued in 2021 with the EMA-FDA-HC Big Data Cluster working towards a RWE collaboration roadmap, exchange of experience with federated data networks to inform the design of DARWIN EU ® and collaboration on COVID-19 observational studies through ICMRA.

In 2022, the data standardisation strategy will be updated with Veterinary requirements, the BDSG & EUNDB will make proposals for actions needed to implement the recommendations of the data standardisation strategy. A real-world evidence workshop with international regulators is planned for

mid-2022. Development of an ICH pharmacoepidemiology safety studies guideline has also been planned.

There was strong support from stakeholders for establishing international collaboration in big data is a foundational activity that will benefit all downstream activities.

Key recommendations from panellists and stakeholders:

- FDA noted **international collaboration/convergence is beneficial**, and we need to continue to share guidance in preparation, work on a collaboration road map, discuss use cases and share lessons learned.

- Health Canada supported moving towards **more alignment and harmonisation**, **not only amongst regulators but also research teams across regions**, especially in the context of the current global pandemic.

- PMDA underlined that all the pre-requisites needed to use and understand RWE are known but **standards are lacking internationally** and might influence the quality of the regulatory use and affect the trust into the regulatory process. **Standards would drive global collaboration**.

- EFPIA representative also called for **international convergence, not only on standards, guidance and best practices but also on terminologies and systems**.

Concluding remarks

The Forum was closed by Marco Greco (EPF) and Peter Arlett (EMA) reminding stakeholders of the key events and consultations planned for 2022.

Collaboration has been a recurrent theme in this Big data Forum and as such it has continued to enable multi-stakeholder inputs in the work of the BDSG. By continuing to work together in the same direction, by 2025, we aimed to achieve the data driven transformation of the EU regulatory network to better support the development of innovated medicines, deliver life-saving treatments to patients more quickly and optimise the safe and effective use of medicines for EU patients.