13 December 2024
EMA/471579/2024

Big Data Steering Group (BDSG): 2024 report

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000

CONTENTS

1. Introduction

The Phase II report of the HMA-EMA joint Big Data Task Force (BDTF) and the proposal to establish a joint BDSG (superseding the Big Data Task Force) were endorsed by the Heads of Medicines Agencies (HMA) in November 2019 and EMA Management Board (EMA MB) in December 2019. The BDSG was established to advise HMA and EMA MB on the recommendations of the Big Data Task Force, covering human and veterinary medicines. The full mandate of BDSG can be found here.

In 2024, significant progress continued to be seen in the transformation to a more data-driven regulation, in line with the Network Strategy to 2025 (in particular for the strategic theme 2 'Data analytics, digital tools and digital transformation'), the BDSG workplan, and the multi-annual AI workplan.

The fourth BDSG workplan was published in July 2023 to progress the activities launched in 2020 and to address new topics. The Multi-annual AI workplan was published in December 2023. The midterm report of the Network Strategy to 2025 published in 2024 highlighted the good progress made during critical period of the pandemic. In 2024 the BDSG met 11 times virtually.

This report provides a summary of the key activities and achievements of the BDSG in 2024 and includes the AI workplan delivery which falls under the BDSG's responsibilities.

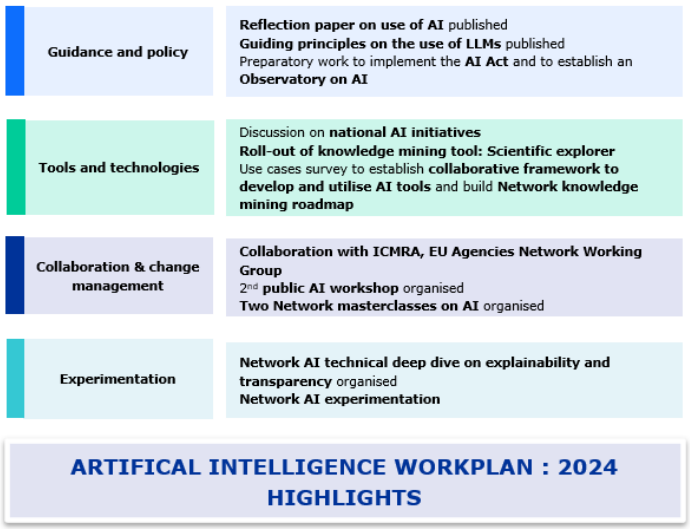
"By delivering a regulatory system able to integrate Big Data into its assessment and decision making, we can support the development of innovative treatments more quickly and optimise the safe and effective use of medicines." **Big Data Task Force Final Report, January 2020**[1]

2. 2024 highlights

Figure 1 below provides a summary of the key BDSG highlights presented in the context of the priority recommendations of the BDSG workplan.

[1] HMA-EMA Joint Big Data Taskforce Phase II report: Evolving Data-Driven Regulation

Figure 2 below provides a summary of the key BDSG highlights presented in the context of the AI workplan.



3. Description of the 2024 highlights of the BDSG workplan

3.1. Recommendation 1: DARWIN EU®

DARWIN EU was launched in February 2022 with the establishment of the DARWIN EU ® coordination centre. DARWIN EU has continued to grow its real-world data (RWD) network throughout 2024. DARWIN EU is now instrumental in enabling EU network capacities to deliver valid and reliable evidence based on RWD studies to complement information from clinical trials and other sources of evidence and support regulatory decision making.

DARWIN EU is now operational and can support more and more evaluation work of EMA's scientific committees and the national competent authorities. It can also support organisations such as the European Centre for Disease Prevention and Control, Health technology assessment bodies and payers with specific studies.

During 2024, 10 additional data partners have been selected for on-boarding into the network, further expanding the network with additional geographies and data types. In total, 20 data partners are now providing access to 25 different data sources and to data from more 130 million patients from 13 European countries. The list of newly onboarded data partners is published on the DARWIN EU website.

The data partners generate real-world evidence (RWE) from sources such as hospitals, primary care practices, health insurances, registries and biobanks to support regulatory activities of EMA's scientific committees and national regulators in the EU. Study protocols and results of all DARWIN EU studies are publicly available in the new HMA-EMA Catalogue of RWD studies.

DARWIN EU is now the main RWE generation pathway used by EMA to support regulatory decisions. In 2024 (up to end of December), 62 research topics have been requested/offered and 18 studies have been completed via DARWIN EU. Since its launch, a total of 58 studies have been initiated. Of these 29 are completed and 29 studies are ongoing. This includes 13 studies to inform vaccine safety and effectiveness, as well as studies in public health emergencies.

The number of studies will be further scaled up with the yearly onboarding of new data partners up to a total of 40 data partners. The aim is to deliver close to 100 studies per year from 2025 onwards to enhance more data-driven decision making on medicines in the EU.

DARWIN EU continued to be supported by the multi-stakeholder DARWIN EU Advisory Board to provide strategic advice and recommendations on DARWIN EU, its possible integration as a node within the European Health Data Space (EHDS), on how to ensure coordination and alignment with relevant European and EU Member State initiatives and policies, and on how to optimise communication on DARWIN EU with stakeholders. Agendas and minutes are published on the DARWIN EU ® webpage.

As part of its continuous engagement with stakeholders and partners, two DARWIN EU training events were organised in 2024. Recordings are available on the EU Network Training Platform (EU NTC) as e-learning for EMA and EU regulatory network only:

- Darwin EU Spring School 'Analyses conducted through the DARWIN EU Network': to increase the understanding of different types of analyses conducted through the DARWIN EU Network. The training aimed to inform colleagues in the regulatory network, so they are better equipped with the knowledge and tools to support their conduct, review, and interpretation of the results in the context of their work.
- DARWIN EU Autumn School: to give an overview of the analytics used in DARWIN EU including population-level disease epidemiology and cohort survival analysis.

In 2024, DARWIN EU continue to take part in the EHDS2 pilot with a use case on coagulopathy and COVID-19. EMA also continued to engage with the EHDS 'Joint Action' to ensure alignment.

> Access to and analysis of real-world data complements evidence from randomised clinical trials to enable timely and reliable evidence for the development, authorisation and supervision of medicines for patients. **Big Data Task Force Final Report, January 2020**[2]

3.2. Recommendation 2: Data quality and representativeness

The RWD quality chapter of the EU Data Quality Framework for EU medicines regulation has been released for public consultation in November 2024. The RWD chapter offers insights on key considerations for RWD quality including practical recommendations that cover characterisation of the systems and processes that underpin data. It also provides a set of metrics to assess data quality dimensions and a guideline on assessing data quality in relation to a research question.

A chapter on Adverse Drug Reaction (ADR) data quality is in preparation for 2025.

To support the preparation and implementation of the European Health Data Space (EHDS), BDSG also collaborated in 2024 with the QUANTUM project for HealthData@EU. The project aims at addressing the growing need for developing a data quality and utility framework to describe the characteristics and the potential usefulness of datasets that exist across Europe from multiple sources, and ultimately design and develop a data quality and utility label for the EHDS.

[2] HMA-EMA Joint Big Data Taskforce Phase II report: Evolving Data-Driven Regulation

By understanding and increasing data quality, the selection of data and interpretation of study results is informed, and the evidentiary value of studies can be judged. **Big Data Task Force Final Report, January 2020**[3]

3.3. Recommendation 3: Data discoverability

HMA-EMA catalogues of real-world data sources and studies

The new HMA-EMA catalogues of real-world data sources and studies were launched in February 2024. The catalogues help medicines regulators, researchers and pharmaceutical companies to identify the most suitable data sources to address specific research questions and support the assessment of study protocols and results. They promote transparency, encourage the use of good practices, and build trust in research based on RWD.

The initiative builds on more than 15 years of operation of the former databases, developed by the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP):

- The catalogue for RWD sources enhances and replaces the ENCePP Resources Database, an EMA-coordinated index of resources of available research organisations, networks and data sources in the fields of pharmacoepidemiology and pharmacovigilance within Europe.
- The catalogue for RWD studies expands and replaces the European Union electronic register of post-authorisation studies (EU PAS Register®).

As part of this initiative, the ENCePP website has been renewed. While some data sources and all centres and networks have migrated to the new catalogues replacing the ENCePP Resource Database, other content, such as ENCePP Guide on Methodological Standards in Pharmacoepidemiology and the ENCePP Code of Conduct, remain available on the ENCePP website renewed in 2024.

The new catalogues introduce various improvements. Aligning with 'FAIR' data principles for Findable, Accessible, Interoperable and Reusable data, they use an agreed set of metadata to describe and connect data sources to studies. This is based on the metadata list for real-world data sources and studies published by the HMA-EMA Big Data Steering group in 2022. In addition, search on a wider set of metadata, enhanced view, export and data submission functionalities have been implemented in the catalogues.

As of October 2024, the RWD Catalogues contain 232 registered data sources and 2,943 studies geographically distributed across various regions of the world. Europe has the highest number of data source and study registrations, with Italy (60 data sources), the United Kingdom (60), Germany (59), and Spain (57) as the top countries leading in data sources. The United Kingdom (1,015 studies), Germany (880) and Spain (776) are the top three for study registrations. Outside of Europe, the most represented countries in data sources are the United States (16), Israel (14) and Australia (11) and for studies - the United States (715), Canada (207) and Japan (134).

Stakeholder engagement continued in 2024 to collect detailed metadata on selected data sources, aiming to populate the catalogues with Real-World Data sources and non-interventional studies. All European data holders, marketing authorisation holders, networks, researchers, and institutions who are interested in having their data used for medicines regulation or are obligated by policy on non-interventional post-authorisation safety studies (PASS), are continuously encouraged to register and keep information up to data in the catalogues.

[3] HMA-EMA Joint Big Data Taskforce Phase II report: Evolving Data-Driven Regulation

These are accompanied in 2025 by the final Good Practice Guide for the use of the Metadata Catalogue of Real-World Data Sources and user guide.

Patient experience data, mobile health data and social media data

Reinforcing patient relevance in evidence generation is a key priority in the EMA's Network Strategy and the realisation of a data-driven regulatory network. In 2024, BDSG discussed the reflection paper on the best EU approach to define, generate, collect and analyse Patient Experience Data (PED), in anticipation of a public consultation in early 2025.

In 2024, to continue to keep abreast with technological and methodological changes, the BDSG discussed the utility of mobile health (mHealth) data and social media data to support regulatory decision-making. BDSG published two expert reviews exploring the utility of mobile health data and social media data, when collected in real-world settings, such as clinical care or the daily life of patients.

Acknowledging challenges in terms of data quality and protection of patient privacy, mHealth tools such as smartphones, health applications, smartwatches and other wearables can generate a large variety of detailed patient data like heart rate and body temperature, and was found to be useful as a complementary source of data for EU medicine regulation in three domains: 1) to support planning and validity of applicant studies, 2) to support the understanding of clinical context, and 3) to investigate associations of products on safety and efficacy outcomes and impacts of regulatory actions.

Despite challenges related to access, quality and ethical use of social media data, various patient data can be extracted from social media, including demographic data, patient experience data (PED), and data related to drug use and disease factors. In some specific cases, such as medicine abuse / misuse or misinformation, social media data might be used as a complementary source of evidence. The BDSG and its successor, the Network Data Steering Group (NDSG) will continue to monitor progress in this area.

Finally, the reports propose a set of points for consideration for future actions to increase the utility of mHealth data and social media data in regulatory decision-making.

Agreement on metadata to describe and identify data sets is enabling data discoverability (including through a publicly available catalogue of real-world datasets) and increasing the ability to judge the evidentiary value of non-interventional (observational) studies and real-world data sources. **Big Data Task Force Final Report, January 2020**[4]

3.4. Recommendation 4: EU Network skills in Big Data

The Big Data Training Signpost was updated in 2024 to provide a helpful resource listing all existing EU NTC trainings in biostatistics, pharmacoepidemiology, data science and pharmacogenomics (Link for NCAs | Link for EMA).

After last year's launch of the Big Data training curricula with the rollout of the first two modules related to the Pharmacoepidemiology and Real-World Evidence (RWE) domain, two additional modules have been rolled out in 2024 on 'Study protocol and reporting' and 'Statistical methods applied to RWE' (Link for NCAs | Link for EMA).

[4] HMA-EMA Joint Big Data Taskforce Phase II report: Evolving Data-Driven Regulation

The *Study protocol and reporting* module covers essential aspects of pharmacoepidemiologic/real-world studies, including research question and study feasibility, study design, statistical analysis plan, quality control for data extraction, transformation and analysis, scientific independence and transparency, and reporting of studies and publication. The *Statistical methods applied to RWE* module covers specific statistical applications for non-interventional studies, combining different real-world data sources and longitudinal analyses.

The RWE curriculum is complemented by the Real-World Academy (RWA), which provides a forum for discussion and exchange for network experts. In 2024 four events took place, discussing a registry-based study on SMA, data quality in RWE, data reliability in DARWIN EU ® and RWD in paediatrics.

The first training modules of the Data Science curriculum have also been made available in 2024. The five modules introduce various data science topics and their application in medicines regulation, including big data, data management and data visualization and Artificial Intelligence.

The Data Science curriculum is complemented by the Digital Academy, offering a collection of learning opportunities on AI and Data literacy (managing, collecting and analysing data) (Link for NCAs | Link for EMA).

Overseen by the Methodology Working Party, all the training modules and event recordings are available to the EU medicines regulatory network's staff through the EU Network Training Platform (EU NTC). Further training modules will be published in 2025. Work is also ongoing on the development of a Genomic data curriculum, its outline was agreed with MWP in December 2024.

Training is supporting the development of an expert workforce able to advise on and interpret big data. **Big Data Task Force Final Report, January 2020**[5]

3.5. Recommendation 5: EU Network processes

Real world evidence

In 2024, EMA has published a summary guide for EU regulators and decision-makers on how the Agency can support the delivery of regulator-lead RWE to aid decision-making. The guide includes descriptions of the available types of RWD studies, optimal resources for addressing research questions and details about the process for requesting these studies.

The second report on conducting studies with real-world data was published in 2024, based on experience gathered between February 2023 and February 2024. Learnings not only from DARWIN EU® studies but also from other pathways used to generate RWD studies, the report highlights the European Medicines Regulatory Network's continued efforts to better integrate RWE into regulatory decisions on medicine development, authorisation and monitoring. It provides in addition a progress update on new studies, expanded data partners, and further recommendations and actions.

During the reporting period, 60 new research topics were identified, of which 38 (63%) were triaged via the DARWIN EU pathway, 16 (27%) via the in-house study pathway, and 6 (10%) via the procured studies pathway. A total of 40 studies were conducted, out of which 22 have been completed while 18 were still ongoing. These studies support an extensive range of decision-makers, including EMA's six scientific committees, EMA's working parties, national competent authorities, as well as the European Centre for Disease Prevention and Control (ECDC), Health Technology Assessment (HTA) and payers' organisations and the European Commission.

[5] HMA-EMA Joint Big Data Taskforce Phase II report: Evolving Data-Driven Regulation

For the first time studies were initiated to inform shortage prevention work of critical human medicines, on herbal substances, for HTA and payer organisations, and to support EMA's geriatric medicines strategy.

The report also includes a list of all new studies considered during the reporting period, an updated portfolio of use cases, and information on a survey on the use of RWD at national level.

To ensure high quality decision making in this rapidly developing environment, the MWP launched in Q3 2024 the public consultation on the reflection paper on use of real-world data in non-interventional studies to generate real-world evidence for regulatory purposes. It is aimed at all stakeholders involved in the planning, conduct and analysis of this type of non-interventional studies, including marketing authorisation holders and applicants. The final reflection paper will be published in 2025.

Patient experience data and genomic data use cases

Use cases for Patient Experience Data (PED) have been discussed in the context of the reflection paper on the best EU approach to define, generate, collect and analyse PED, ahead of the launch of the public consultation in early 2025 (see section 3.3).

Use cases for genomic data have been discussed in the context of the joint EC/HMA/EMA multi-stakeholder workshop on pharmacogenomics held in September 2024 (see section 3.10).

Learning from pilots is informing process improvement, guidance development and the establishment of evidentiary value. Enabling high quality and rapid assessment of medicines will improve decision making throughout the lifecycle of medicines and additionally support greater preparedness for health crisis response. **Big Data Task Force Final Report, January 2020[6]**

3.6. Recommendation 6: EU Network capability to analyse Big Data

Clinical study data pilot

Launched in 2022, the clinical study data pilot (formerly known as raw data pilot) investigates the benefits of visualising and analysing clinical study data to support the scientific assessment of medicinal products. The clinical study data are submitted voluntarily by applicants or marketing authorisation holders as part of their respective applications.

The interim report of the pilot was published in September 2024. It offers preliminary insights into the regulatory benefits of including clinical study data in the marketing authorisation application dossier, potentially leading to earlier opinion and authorisation of medicines. It also examines the operational and technical practicalities for future systematic submission of clinical study data as part of the marketing authorisation dossier. Particular focus is given to analyses supporting the assessment (clinical efficacy and safety but also pharmacokinetics and pharmacodynamics, PK-PD) and related decision-making. Furthermore, the pilot includes analyses to guide the selection of trial sites to verify compliance with Good Clinical Practice (GCP). Experience gained with submission and analysis of patient-level data from clinical studies conducted between September 2022 and December 2023 was included in the interim report.

Based on the initial insights from the pilot including early evidence of the potential for such data analysis to speed up the authorisation of new medicines, EMRN has decided to extend the pilot's

[6] HMA-EMA Joint Big Data Taskforce Phase II report: Evolving Data-Driven Regulation

duration and pilot participation requests from pharmaceutical industry will continue to be accepted until further notice. Guidance for industry has also been updated to reflect the new timelines and interim lesson learned. As of December 2024, nine regulatory procedures have been selected for the pilot.

The pilot continues to be supported by the Network Advisory Group on Raw Data and the Industry Focus Group on Raw Data. Support to MAH has been important, and in 2023 EMA published the application of EMA's transparency principles to the clinical study data proof-of-concept pilot. The Network Community on Raw Data has also been established in 2023.

Non-Clinical raw data

The EMA Proof of Concept (PoC) study to evaluate implementation of the Standard for Exchange of Nonclinical Data (SEND) was launched in 2023. The PoC is looking at how SEND could improve the quality and efficiency of routine assessment but also of procedures where data is complex and requires visualization or independent analysis. The PoC is also looking at how it could be useful for procedures where rapid regulatory input is needed (e.g. rolling review, accelerated procedures, PRIME) and for regulatory science projects or read-across activities to facilitate harmonization, policy or guideline recommendations. The first Marketing Authorisation Application with SEND data was received in September 2024.

Additional analytics initiatives

A survey of the EMRN was conducted in Spring 2024 to explore the type of data currently accessed and available to each NCA, assess the existing demands to analyse data and measure the current computing capabilities, including AI, across the network. The survey focused on data used for regulatory decision making (excluding corporate data) and on technical computing capacities (excluding resourcing) and applied to the Human and Veterinary medicines domains. The results showed an appetite within the EMRN for fostering data access and sharing between agencies and developing analytics capabilities at EU level. The feedback from this survey will inform the development of the new Network Data Steering Group workplan and revised AI workplan in 2025. Some of the recommendations, complementary to the Raw Data pilot report recommendations, will also be considered for the potential drafting of an EMRN data analytics strategy in 2025.

An update on the AI work overseen by the BDSG is provided in section 4.

Demonstrating value of raw data analysis and fostering knowledge and expertise within EU network is key to enable high quality and rapid assessment of medicines. **Big Data Task Force Final Report, January 2020**[7]

3.7. Recommendation 7: Delivery of expert advice

The Methodology Working Party (MWP) was established in 2022 and is now updating its 3-year workplan (3rd version was opened for public consultation in Q3 2024 and adopted in November 2024).

The methodology European Specialist Expert Community (ESEC) has been established with 209 experts from more than 21 European countries including EMA. The ESEC strengthens the MWP, bringing together a broad range of expertise organized in Specialised Interest areas that includes:

- Ai & Data Science: 62 experts

[7] HMA-EMA Joint Big Data Taskforce Phase II report: Evolving Data-Driven Regulation

- Biostatistics: 80 experts
- Clinical Pharmacology: 77 experts
- Modelling & Simulation: 59 experts
- Pharmacogenomics: 28 experts
- Real World Data: 88 experts

The MWP launched in Q3 2024 the public consultation on the reflection paper on use of real-world data in non-interventional studies to generate real-world evidence for regulatory purposes (refer to section 3.5) and publish the reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle in 2024 (refer to section 4.1). The development of guidance on RWE and AI will be initiated in 2025 as part of the delivery of the MWP workplan.

Expert advice including on advanced analytics, real world evidence and 'omics is empowering robust assessment and decision-making by regulatory committees. **Big Data Task Force Final Report, January 2020**[8]

3.8. Recommendation 8: Governance framework

Informed by the input of experts from the Big Data Steering Group (BDSG) and the Network Data Board (NDB), the Network data strategy was launched for public consultation in Q4 2024. The Network strategy aims to provide the vision, principles and objectives for implementing effective data management so that the EMRN can be transformed into a data-driven regulatory ecosystem that promotes innovation, efficiency, and better coordinated decision-making among national authorities and the EMA. They key strategic objectives of the strategy are related to data governance, data quality management, Interoperability, data cataloguing and metadata management, knowledge and change management, and analysis of data.

In 2024, the Network data governance was also reviewed. A decision was taken to unify the two existing data governance groups, the BDSG and the NDB into one network data governance group, the Network Data Steering Group (NDSG). The mandate of the NDSG was endorsed by HMA at its September 2024 meeting and EMA Management Board at its October 2024 meeting.

The NDSG will support the delivery of the EMAN Strategy to 2028 (including Theme 2: Leverage data, digitalization and AI) and of the EMRN Data strategy, both currently under public consultation.

Like the BDSG, the NDSG will be a strategic advisory group established to maximise data interoperability, exchange and use across the EU network, the access to data and generation of evidence, and the beneficial utilisation of AI. The NDSG will make proposals and give advice to HMA and EMA Management Board for the prioritisation, planning and monitoring of actions relevant to the EMAN strategy to 2028 (particularly its Theme 2: Leverage Data, Digitalisation and AI) and the EMA multi-annual workplan. Its scope will encompass regulatory data, including supply data, and data supporting evidence on medicines.

Building on the model of the BDSG, its mandate will also include stakeholders' engagement and network capability/capacity activities. It will continue to be co-chaired between EMA and HMA and its members will represent NCAs, EMA, EC as well as key stakeholders (however interaction with Industry

[8] HMA-EMA Joint Big Data Taskforce Phase II report: Evolving Data-Driven Regulation

will be via regular bilateral meetings). The group will be appointed until end of 2028, in line with EMANS to 2028.

The BDSG continued to prepare for a changing policy environment with the future European Health Data Space (EHDS) and the revised Pharmaceutical Strategy for Europe via regular updates from the European Commission and participation to various fora and workshops.

Secure and ethical data governance is an enabler for secondary use of healthcare data and the work of the BDSG seeks to support stakeholders to navigate and comply. **Big Data Task Force Final Report, January 2020**[9]

3.9. Recommendation 9: International initiatives

Progress in convergence with international partners on RWD/RWE continued in 2024 and contributed to establish the value of RWE and enabling its use.

In June 2024, following a 3-month public consultation, the Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) Assembly adopted a reflection paper called "Pursuing Opportunities for Harmonization in Using Real-World Data to Generate Real-World Evidence, with a focus on Effectiveness of Medicines".

In November 2024 the related new topic proposal entitled "Considerations for the Use of Real-World Evidence (RWE) to Inform Regulatory Decision Making with a focus on Effectiveness of Medicines" was sent to ICH. The aim of this proposal is to harmonise RWD/RWE terminologies, metadata and establish general principles for assessing studies using RWD to support regulatory decision-making worldwide.

The reflection paper as well as the proposal was co-developed by EMA, FDA and Health Canada following a joint statement by ICMRA calling for international collaboration in generating and using real-word evidence in regulatory decision making. This marks an important commitment from ICH towards harmonisation of regulatory RWE guidance.

In 2024, the public consultation on the ICH M14 draft guideline on general principles on plan, design and analysis of pharmaco-epidemiological studies that utilize real-world data for safety assessment of medicines was launched. The draft guideline provides internationally harmonised guidance, recommendations and best practices for the conduct of pharmaco-epidemiological studies using real-world data for the assessment of the safety of medicines. The aim is to streamline the development and regulatory assessment of study protocols and reports and make them more acceptable to health authorities worldwide. An overview of the comments received is published on EMA website.

In addition, the former "COVID-19 Real-World Evidence Observational Studies Working Group (WG)" published in November the paper Collaborative Real-World Evidence Among Regulators: Lessons and Perspectives reflecting on 3 years of collaboration. The WG has been repurposed as "Real-World Evidence for Public Health Emergencies (PHE) Working Group". It is co-chaired by EMA and Health Canada. The first meeting took place in July 2024 with fifteen regulatory agencies, and the second meeting took place in December 2024. The WG aims to be a global forum for regulators to enhance the efficiency and effectiveness of responses to PHEs by conducting collaborative RWD studies. The members are currently working on the development of general principles and operational aspects for conducting collaborative studies, as well as on a list of potential topics of interest to test the system.

[9] HMA-EMA Joint Big Data Taskforce Phase II report: Evolving Data-Driven Regulation

The digital transformation of clinical trial information is currently being strengthened with initiatives such as the digital protocol (ICH M11) or the Patient Level Data Project. One critical aspect of this transformation is the need for global standard for passing data between systems. Fast Healthcare Interoperability Resources (FHIR) is currently being developed as a free and open source to enable the data transfer of protocol data and can be applied to summary results in the domain of interventional trials. Learnings from this initiative will likely benefit the technical solutions for non-interventional studies.

Finally, in 2024, BDSG discussed the ICH M15 guideline on general principles for model-informed drug development. This guideline defines an harmonised framework for evaluating model-informed drug development (MIDD) evidence and provides recommendations for related planning and regulatory interactions, implementation, reporting, and submission. The guideline is under public consultation until February 2025.

Convergence with international partners on standards and guidelines will leverage best expertise and minimise burden on stakeholders. **Big Data Task Force Final Report, January 2020**[10]

3.10. Recommendation 10: Stakeholder engagement

Collaborations with external stakeholders and partners continued in 2024 and informed the work of the BDSG.

BDSG supported the organisation of the joint HMA/EMA ACT EU Clinical Trials Analytics Workshop in January 2024 to discuss use cases for analysis of data about clinical trials in the EU through clinical trial registries (CTIS and EudraCT) and to help guide EU decision-makers on improving access and usability of clinical trials data.

The joint HMA/EMA multistakeholder workshop on patient registries was held in February 2024 to discuss the EMA qualification procedure for patient registries (with the aim to clarify the benefits, identify current limitations, and propose measures to optimise the process) and establish the value and enable the use of patient registries for regulatory decision-making by considering contexts of use for which registry data are 'fit for purpose', and examining tools to support data discoverability and assessment. Outcomes linked to qualification of registries will be implemented through the action plan for the future proofing Qualification of Novel Methodolgies published in September 2024.

The joint HMA/EMA multi-stakeholder workshop on data quality framework for adverse drug reaction reporting was held in March 2024 to bring together experts in the field to build on their extensive experience and knowledge relating to ADR data quality, and inform the drafting of the next chapter of the data quality framework for EU medicine regulation.

The joint HMA/EMA Big Data Steering Group workshop on real-world evidence (RWE) methods was held in June 2024 to collect views of stakeholders and experts on the draft RWE reflection paper and on priorities for further regulatory guidance development and collaboration beyond the reflection paper.

The joint EC/HMA/EMA multi-stakeholder workshop on pharmacogenomics was held in September 2024 to identify priority areas for additional regulatory action to promote the clinical implementation of pharmacogenomics, discuss how medicines regulators can facilitate the uptake of genomics by national healthcare systems, how to leverage genomic data linked to real-world data sources with examples of

[10] HMA-EMA Joint Big Data Taskforce Phase II report: Evolving Data-Driven Regulation

current studies using such data, and how to increase the regulatory impact of EU-funded projects in personalised medicine. The outcome of the workshop will inform a roadmap towards the clinical implementation of pharmacogenomics in Europe.

The joint HMA/EMA multi-stakeholder workshop on artificial intelligence (AI) - enabling the safe and responsible use of AI was held in November 2024 to update stakeholders on the state-of-the-art of AI developments, policy and legislative environment and the revision of the AI reflection paper, and HMA/EMA activities conducted under the multi-annual AI workplan. The outcome of the workshop will inform the revision of the multi-annual AI workplan and the drafting of AI network research priorities roadmap.

The fifth annual joint HMA/EMA Big Data Stakeholder Forum was held in November 2024 to discuss progress on evidence generation, interoperability, use and exchange of medicines data across the EU network, and inform the future work planning of the group.

The Methodology Working Party Interested Parties meeting took place on June 2024, including industry to give stakeholders the opportunity to raise and share reflections on relevant methodological topics and for MWP to identify gaps in its workplan with the aim of prioritising guideline development accordingly.

Two dedicated BDSG industry stakeholder meetings (in May and November 2024) were organised in the context of EMA's continuous endeavours to foster regular interactions with industry stakeholders on topics of common interest. Agenda and minutes are published on EMA website. A specific Industry focus group on RWE was established in 2024 to share knowledge and experience with use of RWD and generation of RWE to advance integration of relevant and reliable RWE in regulatory decision making.

The EMA Big data newsletter, published 4 times in 2024, has provided an update on progress in implementing the workplan of the HMA-EMA Big Data Steering Group.

The BDSG is assessing communication and engagement efforts with partners and stakeholders via a survey launched in September 2024.

The BDSG agreed the change management strategy that includes the key changes to a more data-driven medicines regulation, the key stakeholder groups and the approach to engage them, the change management activities delivery plan, and the actors responsible for executing the strategy. The Change management activities delivery plan was published in July 2024.

> Listening to stakeholders and leveraging their work is optimising and maximising transformation to data-driven regulation. Transparency and communication build trust in what is delivered. **Big Data Task Force Final Report, January 2020**[11]

3.11. Recommendation 11: Veterinary recommendations

The last 12 months have seen great progress in the implementation of the European Veterinary Big Data strategy 2022- 2027. The overall strategy implementation is on schedule, and delivery highlights include:

[11] HMA-EMA Joint Big Data Taskforce Phase II report: Evolving Data-Driven Regulation

- Completion of the 'Big Data in Veterinary Medicines Regulation: A Data Landscape Analysis' study which mapped the veterinary data landscape and identified veterinary real world data sources.
- The go-live of Data Literacy training on the EUNTC platform (Link for NCAs | Link for EMA).
- The Union Product Database (UPD) Data Quality Framework has been in production for 1 year, reducing the number of UPD data quality issues by 40%.

The next iteration of the EU Veterinary Big Data Workplan to 2022-2025, which details work items from 2025 to 2027 is being developed in partnership with the Veterinary Data Hub. The Veterinary data hub is a multi-disciplinary team of Network experts from Belgium, France, Germany, Portugal, Spain and Sweden. Industry stakeholders have also contributed thoughts on the EU Veterinary Big Data workplan scope.

The fourth Veterinary big data stakeholder forum was held in October 2024. The agenda covered several topics of focus in the big data domain. The event highlighted the transformational opportunities in the lifecycle of veterinary medicinal product regulation enabled by big data. There were active contributions from industry, academia and the regulatory network with presentations on the Veterinary Big Data strategy implementation, use of AI by the Industry and a horizon scan of themes arising in the big data and AI domain.

Synergies exist in the use of data between the human and veterinary domains that can catalyse our transformation. **Big Data Task Force Final Report, January 2020**[12]

4. Description of the 2024 highlights of the AI workplan

4.1. Guidance, policy and product support

Guiding principles on the use of large language models (LLM) in regulatory science and for medicines regulatory activities were published in September 2024 and are now available for the Network staff on how to use large language models in their work. They can help staff in areas such as processing extensive documentation, automating data mining and optimising routine administrative tasks. They also present challenges such as providing irrelevant or inaccurate responses and posing potential data security risks. They focus on LLM for text generation. This document aims to promote the safe, responsible and effective use of this category of AI technology. The guiding principles are a living document that will be regularly updated.

This has been supported by an information campaign that included the factsheet: Four principles for safe and responsible use of large language models and a masterclass (see section 4.3)

Following the public consultation, the final reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle was published on 30 September 2024. This paper reflects on principles relevant to the application of AI and machine learning (ML) at any step of a medicines' lifecycle, from drug discovery to the post-authorisation setting.

Preparatory work to implement the AI Act and to establish an Observatory, that includes horizon scanning on AI, has started in 2024.

[12] HMA-EMA Joint Big Data Taskforce Phase II report: Evolving Data-Driven Regulation

4.2. Tools and technologies

In 2024, BDSG continued to be a place to discuss national AI initiatives in particular two national initiatives from Sweeden (AI@MPA) and Belgium (ERAMET project).

Launched in March 2024, Scientific Explorer is an AI enabled knowledge mining solution for EU regulators that facilitate easy, focused and precise searches of regulatory scientific information from network's sources. It supports scientific decision-making by providing access to relevant scientific information. The first release is dedicated to scientific advice procedures for human medicines.

The AI@MPA tool is a web-based service provided by the Swedish Medicines Agency to the European Medicines Agency (EMA) and other national competent authorities (NCA). It uses artificial intelligence to help assessors to search and identify relevant guidelines, EPARs (European Public Assessment Reports), qualification opinions and SmPC (Summary of Product Characteristics).

The survey for collecting feedback from EMRN on current and desired analytics capabilities is described in section 3.6 of this document.

A survey to the EU regulatory network has been launched in December 2024 to identify potential use cases and national solutions/initiatives on AI, that have already been implemented or planned to be implemented, and for which their use could be extended to the entire regulatory network. The results of the survey will aid to define a collaborative framework across the network to develop and utilise AI tools for enhancing medicines evaluation processes. It will also inform the development of the Knowledge Mining network's roadmap to be completed in 2025.

4.3. Collaboration and change management

Collaboration with stakeholders and partners continued in 2024 and international level and under the umbrella of the International Coalition of Medicines Regulatory Authorities (ICMRA). At European Union level, EMA has started chairing the EU Agencies Network Working Group (EUAN WG) on AI. The WG aims to support the agencies network on the implementation of AI fostering knowledge sharing and increasing the AI maturity level to provide guidance and common approaches. As main activity during the year, a physical workshop with all the members of the WG was held in September 2024 at EMA premises, onboarding new agencies and helping to shape a workplan for the coming years. The event primarily focused on knowledge-sharing sessions and included various exercises aimed at collaboratively preparing the future Community/WG AI workplan. It featured updates on current AI use cases, presentations from community members, and inspiring talks from external speakers. The event showcased the strong partnerships and unique capabilities within the Community/WG, highlighting the importance of collaboration among members, with all participants united in their efforts to shape the community’s future.

The European Specialised Expert Community (ESEC) of the EMA Methodology Working Party (MWP) is now fully established and the 62 members of the Special Interest Area (SIA) on AI started to work together, meeting virtually for the 1st time in 2024. It now provides a Network community of practice as a forum for collaboration and knowledge sharing.

In addition to the rollout of the Data Science curriculum in 2024 (see section 3.4 of this document), two AI Masterclasses on the use of Large Language Models (I) and Hands on with the guiding principles on the use of LLMs (II) were delivered for the EU network in 2024. Recording of both masterclasses are available on the EUNTC platform.

The joint HMA/EMA multi-stakeholder workshop on artificial intelligence (AI) - enabling the safe and responsible use of AI was held in November 2024. This event was attended by approximately 1700 online participants and by around 70 in person attendees (See also section 3.10 of this document). The outcome of the workshop will inform the revision of the multi-annual AI workplan and the drafting of AI network research priorities roadmap.

4.4. Experimentation

The first AI Technical Deep Dive for the network on explainability and transparency took place in October 2024.

Informed by the discussion at the joint HMA/EMA multi-stakeholder workshop on artificial intelligence (AI) - enabling the safe and responsible use of AI, the drafting of the AI network research priorities roadmap will start in 2025, and experimentation cycles will be launched in line with those priorities.