20 February 2022
EMA/40400/2022

Big Data Steering Group (BDSG): 2021 report

This report provides a summary of the key activities and achievements of the BDSG in 2021. Significant progress in the transformation to data-driven regulation continued in 2021, in line with the Network Strategy to 2025 and BDSG workplan.

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.

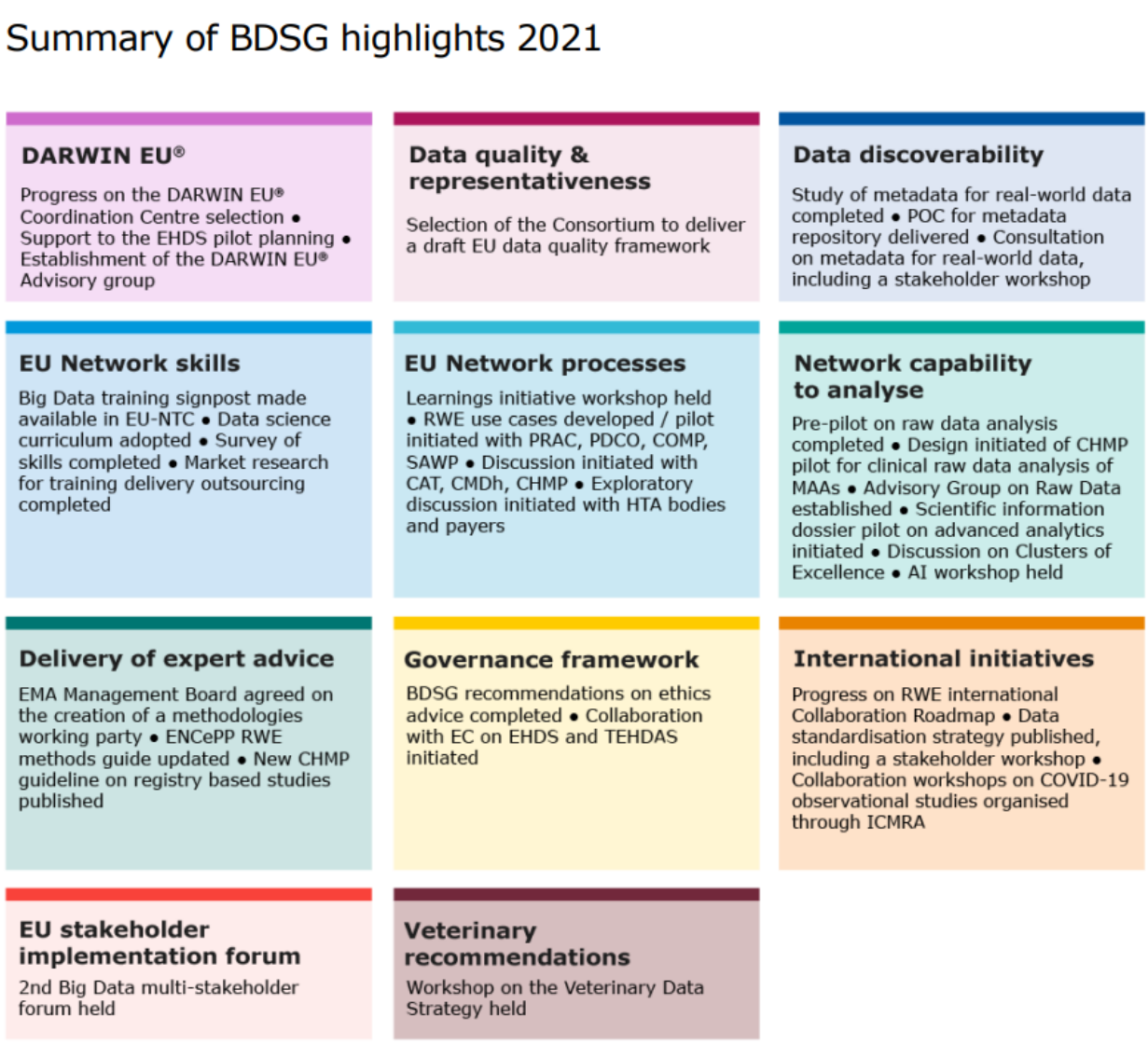
The BDSG adopted its second workplan in June 2021, setting its priorities for 2021 and 2022.

In 2021 the BDSG met 11 times virtually and progress was made in line with the agreed workplan.

"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**

2021 highlights for the priority recommendations

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



Description of 2021 Highlights

Recommendation 1: deliver a sustainable platform to access and analyse healthcare data from across the EU (Data Analysis and Real-World Interrogation Network – DARWIN EU®)

DARWIN EU ® has seen significant progress in 2021, moving from a vision to launch the tender to select the coordination centre that will operate the DARWIN EU ® network on behalf of the EU regulatory network.

The tender procedure was launched in June 2021 and concluded in February 2022 with the appointment of Erasmus University Medical Center Rotterdam as the DARWIN EU coordination centre. This marks the launch of the establishment of DARWIN EU ®.

Including key stakeholder representatives, the DARWIN EU ® Advisory Board was formed, and provided 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 supported communication on DARWIN EU ®. Three Advisory Board meeting took place in 2021 and the agenda and minutes are published on the DARWIN EU ® webpage.

The BDTF recommendations foresaw EMA fees as the source of long-term funding for the operational phase of DARWIN EU ® and EMA continued to work with the European Commission (EC) services to support the impact assessment for a revised EMA fees regulation to ensure sustainable funding of DARWIN EU ® at no direct cost to the public purse.

As DARWIN EU ® will act as a pathfinder initiative and use case for the proposed European Health Data Space (EHDS), EMA actively collaborated with EC and submitted a use case for the pilot planned to explore the potential for the EHDS. EMA also continued to engage with TEHDAS and its work packages to ensure alignment.

Access to and analysis of real-world data will complement evidence from randomised clinical trials and support the development, authorisation and supervision of medicines for patients.

Recommendation 2: Establish an EU framework for data quality and representativeness

In late 2021, EMA appointed the consortium that will deliver the draft of a first EU data quality framework for data used for regulatory purpose. Informed by the analysis of existing data quality initiatives and discussion with a wide range of stakeholders via a workshop, the EU data quality framework will provide high level principles, applicable across big data and data sets used in regulatory submissions and decision making, and a series of applied use-cases for regulatory purposes (e.g. data quality measures for healthcare data)

Collaboration with TEHDAS on data quality was established and will intensify in 2022.

By understanding data quality, the selection of data and interpretation of study results is informed, and the evidentiary value of studies can be judged.

Recommendation 3: Enable data discoverability

Informed by consultation with stakeholders (including a dedicated workshop and survey with ENCePP) the first list of real-world metadata for regulatory purpose was created in 2021 and included the

definitions of key concept and the first criteria for inclusion of databases in the proposed real-world metadata catalogue (i.e. a data source catalogue).

Building on the real-world metadata list, the proof-of-concept for the real-world metadata catalogue was successfully implemented in EMA architecture and will serve as the basis for the delivery of the definitive EU regulatory real-world metadata catalogue and the study catalogue replacing the EU PAS register. The work to build the public interface to access the catalogue via EMA website and start populating the catalogue has also been initiated in 2021.

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

Agreement on metadata to describe and identify data sets will enable data discoverability including through a publicly available catalogue of real-world datasets.

Recommendation 4: Develop EU Network skills in Big Data

In 2020, the 'Big Data Training Signpost' tool for staff in the National Competent Authorities and EMA was delivered and the Biostatistics and Clinical Trial Methodology and Pharmacoepidemiology/real-world evidence curricula was adopted. In 2021, an additional Data Science curriculum was adopted by the BDSG and a survey of EU Network skills was completed to inform on future needs and prioritisation. The BDSG will act as the Data Science curriculum owner and is supported by the Big Data Curriculum Steering Group.

Market research for potential outsourcing of training delivery was completed and will allow to launch a procurement in 2022 to select external training providers that will deliver content for these three training curricula. Ultimately these trainings will be rolled out to the EU Network via the EU Network Training Centre.

Training will support the development of an expert workforce able to advise on and interpret big data.

Recommendation 5: Strengthen EU Network processes for Big Data submissions

In 2021, PRAC established routine processes for RWD access to 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 (1 study), the Paediatric Committee (2 studies) and the Orphan Committee (1 study). Preliminary discussion on possible pilots with the Committee on Advanced Therapies, the Coordination Group for Mutual Recognition and Decentralised Procedures – Human, the Committee for Medicinal Products for Human Use as well as Health Technology Assessment bodies and Payers were organized in 2021.

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). The summary report of the forum can be accessed here. This workshop will further inform process and guidance improvement in 2022-2023.

Finally, the BDSG continued to learn from recent EMA experience with medicinal products in combination with digital health / medical devices as part of the MAAs, variations or as a printed QR

code/URL on package leaflets and outer cartons. Most of the EMA interactions with such medicinal products occur through either Scientific Advice or the EMA's Innovation Task Force meetings, with the focus on data collection methods for future analysis to support MAAs. Recommendations to strengthen the process to advise on data qualification will be discussed in 2022.

Learning from data in regulatory submission will inform process improvement, guidance development and the clarification of evidentiary value.

Recommendation 6: Build EU Network capability to analyse Big Data

In 2021, the lessons learned from the clinical trial raw data pre-pilot was finalised and the Advisory Group on Raw Data was established to inform the full pilot that will explore the business case and practicalities for analysis clinical trials raw data in MAAs in 2022.

As a complement to 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 has explored 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). Led by DKMA, the drafting of a discussion paper on Clusters of Excellence has been initiated and has brought together the expertise of DKMA, BfArM, PEI, AEMPS, Infarmed, SMPA and MEB.

The Joint HMA/EMA workshop on artificial intelligence (AI) in medicines regulation was organised in 2021 to inform on the state-of-the art use of AI and Machine Learning in medicines development and hear the views of stakeholders and experts on potential AI actions and to identify any gaps. Developing a framework to access and validate AI and a framework that supports the development of guidelines were identified as priorities and the work to draft guidance on AI in medicines regulation will be initiated in 2022. The summary report of the forum can be accessed here.

Technology is a key enabler to access and analyse big data and its informed selection and use will accelerate transformation to fully data-driven regulation.

Recommendation 7: Modernise the delivery of expert advice

In 2021, the EMA Management Board agreed to the creation of a Methodology Working Party to modernise and strengthen the delivery of expert advice on methodologies bringing together a broad range of expertise (i.e. Big Data, biostatistics, real world evidence, advanced analytics, PK/PD, extrapolation, modelling and simulation, GCP and omics).

Methodology guidelines have continued to be rolled-out including a substantial revision of the ENCePP guide on pharmacoepidemiological methods published in July 2021 and the new CHMP Guideline on Registry based studies in October 2021.

Expert advice including on advanced analytics, real world evidence and 'omics will empower assessment and decision-making by regulatory committees.

Recommendation 8: Ensure data are managed and analysed within a secure and ethical governance framework

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 was progressed and will likely be published in 2022, subject to the publication of the anticipated guidance from the European Data Protection Board.

In addition, BDSG held exploratory discussions on data ethics and prioritised to engage with ongoing initiatives.

The BDSG continued to prepare for the future European Health Data Space (EHDS) and the revised Pharmaceutical Strategy for Europe via regular updates from the European Commission. Upon request of the European Commission, the drafted of the concept paper on legislative change for real-world data including registries was initiated with the participation of members of the BDSG.

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.

Recommendation 9: Collaborate with international initiatives on Big Data

Following extensive stakeholder consultation in 2021, the European Medicines Regulatory Network Data Standardisation Strategy was endorsed by Heads of Medicines Agencies and EMA Management Board. This is a critical element for realising the full potential of data in driving regulatory decisions and it will serve to support the development of globally applicable standards for the human and veterinary regulatory domains.

International collaboration on RWE has continued in 2021 with the EMA-FDA-HC Big Data Cluster working towards a RWE collaboration roadmap and the planning of a Regulatory Workshop in 2022. Exchange of experience with federated data networks in 2021 helped in the design of DARWIN EU ® and collaboration on COVID-19 observational studies took place through ICMRA.

Convergence with international partners on standards and guidelines will leverage best expertise and will minimise burden on stakeholders.

Recommendation 10: Create an EU Big Data 'stakeholder implementation forum'

The second Big Data Stakeholder Forum was held virtually in December 2021. The forum is the opportunity for stakeholders to provide their feedback regarding the Big Data priority recommendations and their perspectives on implementation. One hundred and sixty-five registered participants attended the Forum by webinar in addition to the live stream on the web. Data quality and representativeness, DARWIN EU ® and regulatory processes were the top priorities for 2022 as indicated by the big data forum participants. The summary report of the forum can be accessed here.

Listening to stakeholders and leveraging their work will optimise and maximise transformation to data-driven regulation.

Recommendation 11: Veterinary recommendations

The 2021 Veterinary Big Data Stakeholder forum was the first opportunity to bring together regulators, the pharmaceutical industry, farm management system providers, academia, consumers and practitioners, to build awareness on the use of innovative digital technologies in the veterinary regulatory environment, share needs, ambitions and opportunities, and inspire future activities shaping the development of the European Veterinary Big Data Strategy expected in 2022.

The summary report of the forum can be accessed here.

Synergies exist in the use of data between the human and veterinary domains that can catalyse our transformation.