HMA/EMA Joint Task Force on big data - Survey for National Competent Authorities

Fields marked with * are mandatory.

Background

This survey forms part of the work of the Heads of Medicines Agencies/European Medicines Agency (HMA /EMA) Joint Task Force on big data whose mandate[1] is to explore a number of issues regarding the emerging challenges presented by big data in the regulation of medicinal products for human use.
The term "big data" refers to extremely large sets of information, which require specialised computational tools to enable their analysis and exploitation.
These data might come from:
- data derived from clinical practice: e.g. electronic medical records, claims data, registries;
- “omics” data: e.g. genomics, proteomics, metabolomics;
- social media, m-health and other innovative and novel datasets;
- clinical trials;
- reports on spontaneous adverse reactions.

In the context of Medicines Regulation, the revolution in data and associated analytical tools could mean the submission of data in large amounts or of a complex nature supplementing more traditional analysed and structured data. Alternatively, it could exist as data lying underneath the regulatory submissions, for which it would be crucial to understand their presence and the robustness by which they were generated in order to make a competent evaluation of the submission as a whole.
While the evidence derived from big data sets has the potential to provide significant additional insights for the assessment of the benefit-risk of medicinal products over their entire life cycle, it is recognised there are also substantial challenges in the use of these data.

This interactive e-survey is divided into 3 sections and aims to understand:
- the current state of expertise and experience of National Competent Authorities (NCAs) regarding big data;
- the current challenges identified by NCAs;
- the expectations of and future challenges anticipated by NCAs.

The results are important to develop a future big data strategy and identify the future needs of the European Medicines Regulatory Network in this context. A synopsis of the results will be available as a public output of the HMA/EMA Joint Task Force.

[1] HMA/EMA Joint Task Force on big data Mandate: http://www.ema.europa.eu/docs/en_GB/document_library/Other/2017/03 /WC500224262.pdf and http://www.hma.eu/506.html

How to complete the survey

The survey should be completed as fully as possible by the **9th October 2017**. It is customised to the respondent's profile and can only be submitted electronically. Any question marked with an asterisk (*) require an answer in order to progress through the survey. Free text boxes are allowed for further comments. If you have any questions, you should contact the Secretariat of the HMA/EMA Joint Task Force on big data directly at tse@dkma.dk.

This survey aims to collect only aggregated and anonymous data. It does not require you to disclose your personal identity, although your NCA will be recorded. We will collect and analyse the responses provided and use the analyses in public documents but we will not link them to any other personal information you may provide. Collected data will be kept one year and then destroyed. For security reasons, the survey platform will record IP addresses for every server request accessing the questionnaire.

Current state of expertise

*** 1. Please select the country of your NCA:**

- Austria
- Belgium
- Bulgaria
- Croatia
- Cyprus
- Czech Republic
- Denmark
- Estonia
- Finland
- France
- Germany
- Greece
- Hungary
- Iceland
- Ireland
- Italy
- Latvia
- Liechtenstein
- Lithuania
- Luxembourg
- Malta
- Netherlands
- Norway
- Poland
- Portugal
- Romania
- Slovak Republic
- Slovenia

- Spain
- Sweden
- United Kingdom

2. What is your NCA's current in-house level of expertise relevant to big data analytics? Please indicate the approximate numbers of Full Time Equivalents (FTEs) with expertise in the following analytical areas:

	0	1 to 5	6 to 10	More than 10
* Biostatistics/statistics				
* Bioinformatics				
* Data modelling and prediction				
* Data standardisation				
* Data privacy experts/data anonymisation				
* Machine learning/data mining				
* Natural language processing				
* Data programming				
* IT expertise				
Other areas *(please specify in the comments field below)*				

Any comments

250 character(s) maximum

*** 3. Does your NCA consider it necessary to increase the number of resources in house with expertise to analyse big data?**

- Yes *(please complete the table below)*
- No *(please provide a justification in the comments box below)*

If YES, please indicate with what timescales for each area:

	No concrete timescale	< 1 year	1 to 3 years	> 3 years
* Biostatistics/statistics				
* Bioinformatics				
* Data modelling and prediction				
* Data standardisation				

* Data privacy experts/data anonymisation				
* Machine learning/data mining				
* Natural language processing				
* Data programming				
* IT expertise				
* Other areas *(please specify in the comments field below)*				

* Any comments

250 character(s) maximum

Current state of experience

*** 4. Does your NCA have direct access to external big data sets to inform your decision making process?**

- Yes
- No

* If YES, please select the relevant areas below:

- Genomics (e.g. European Nucleotide Archive (ENA), GenBank)
- Proteomics (e.g. PRIDE Database, Global Proteome Machine Organisation Database)
- Other –omics (e.g. lipidomics (like Lipid Maps), metabolomics)
- Clinical trial datasets
- Adverse Drug Reaction data (excluding Eudravigilance)
- Real World Data/Observational data (e.g. registries, electronic health records etc.)
- Social media data (e.g. twitter data)
- Other *(please specify in the comments field below)*

Any comments

250 character(s) maximum

*** 5. Does your NCA have in-house systems / tools to analyze primary data sets in order to assess the quality of submitted data sets?**

- Yes
- No

* If YES, please select the relevant tools below:

- SAS

- [ ] STATA
- [ ] Hadoop/Spark
- [ ] Spotfire
- [ ] MapR
- [ ] SQL/noSQL (not only SQL)
- [ ] R
- [ ] Python
- [ ] Other *(please specify in the comments field below)*

Any comments

250 character(s) maximum

***6. Does your NCA have any ongoing collaboration with academic institutions within big data?**

- ( ) Yes
- ( ) No

*If YES, please specify which academic institutions for what purpose and which data:

250 character(s) maximum

***7. Has your NCA outsourced any big data-related services?**

- ( ) Yes
- ( ) No

*If YES, please provide details of outsourcing:

250 character(s) maximum

Any comments

250 character(s) maximum

***8. Has your NCA received any requests from any sponsors asking for scientific advice at national level on the applicability or the analysis of big data sets?**

- ( ) Yes
- ( ) No

*If YES, please indicate the approximate number of requests:

- ( ) 1-5
- ( ) 5-10
- ( ) 10-15

- ○ >15

* If YES, please indicate in which areas and specify their source in the comments fields below:

- ☐ Genomics
- ☐ Proteomics
- ☐ Other –omics (e.g. metabolomics, lipidomics)
- ☐ Clinical trial datasets
- ☐ Adverse Drug Reaction data
- ☐ Real World Data/Observational data (e.g. registries, electronic health records etc.)
- ☐ Social media data (e.g. twitter data)
- ☐ Imaging datasets (functional MRI, PET etc.)
- ☐ Other *(please specify in the comments field below)*

Any comments

250 character(s) maximum

9. Please indicate where big data has been used to derive evidence to support regulatory procedures, e.g. to understand disease progression, to better stratify patient populations, to identify an underlying mechanism of action for an ADR?

	Yes	No
* Orphan designation	○	○
* Paediatric Investigation Plans	○	○
* Scientific Advice at national level	○	○
* Marketing Authorisation application	○	○
* Type IA variations	○	○
* Type IB variations	○	○
* Type II variations	○	○
* Annual reassessments	○	○
* 5 years and Annual Renewals	○	○
* PASS / PAES	○	○
* RMP	○	○
* PSUR	○	○
* Referrals	○	○
* Any Other	○	○

Please provide examples of the information submitted and describe the challenges in its asessment:

250 character(s) maximum

***10. Please provide specific examples, if any, where big data sources have been used to derive evidence to support label change decisions:**

250 character(s) maximum

Current challenges

***11. In which regulatory areas does your NCA see big data being first applied? Please select the relevant areas below:**

- ☐ Clinical trials
- ☐ Approval of new medicinal product
- ☐ Confirmation of a conditional MA
- ☐ Change in label of already approved medicinal product
- ☐ Extending an indication
- ☐ Signal detection
- ☐ Signal validation and assessment
- ☐ Validating benefit-risk in high risk populations
- ☐ Other *(please specify in the comments field below)*

Any comments

250 character(s) maximum

12. What are the biggest challenges in the use of big data from the perspective of your NCA? Please prioritise your choices on a scale of 1 to 5 where 1 corresponds to the lowest challenge and 5 the biggest:

	1	2	3	4	5
* Data access	○	○	○	○	○
* Data privacy/legislation on data protection	○	○	○	○	○
* Data security	○	○	○	○	○
* Data quality	○	○	○	○	○
* Data harmonisation accross Europe	○	○	○	○	○
* Integration of multiple data sets	○	○	○	○	○
* Validation of results	○	○	○	○	○

* Data reproducibility					
* Expertise across the regulatory network					
* Challenges of recruiting the relevant expertise					
* Information Technology challenges					
* Other *(please specify in the comments field below)*					

Any comments

250 character(s) maximum

13. How important does your NCA management consider the use of big data to support regulatory decision making? Please indicate a number (1-5), where 5 is very important and 1 is not important:

	1	2	3	4	5
* Now					
* In 5 years' time					

Any comments

250 character(s) maximum

*** 14. Is there any specific national legislation on access to big data in your country?**

- Yes *(please specify in which area e.g. genomics, clinical trial data, healthcare data, social media data etc. in the comments field below)*
- No
- Unknown

* If YES, please specify in which area:

250 character(s) maximum

Any comments

250 character(s) maximum