8 April 2025
EMA/124430/2025
European Medicines Agency

Outcome of public consultation on the Good Practice Guide for the use of the HMA-EMA Catalogues of real-world data sources and studies

Summary report of comments received during the public consultation and next steps

1. Background and consultation

In anticipation of the launch of the HMA-EMA Catalogues of real-world data sources and studies, the draft 'Good Practice Guide for the use of the HMA-EMA Catalogues of RWD sources and studies' (GPG) was developed to provide regulators, researchers, data holders and other interested stakeholders with recommendations on using the catalogues to identify suitable data sources when planning a study. The guide was developed under the HMA-EMA joint Big Data Steering Group lead and is part of the group's 2023-2025 workplan.

EMA invited stakeholders to comment on the draft GPG between 27 September and 16 November 2022.

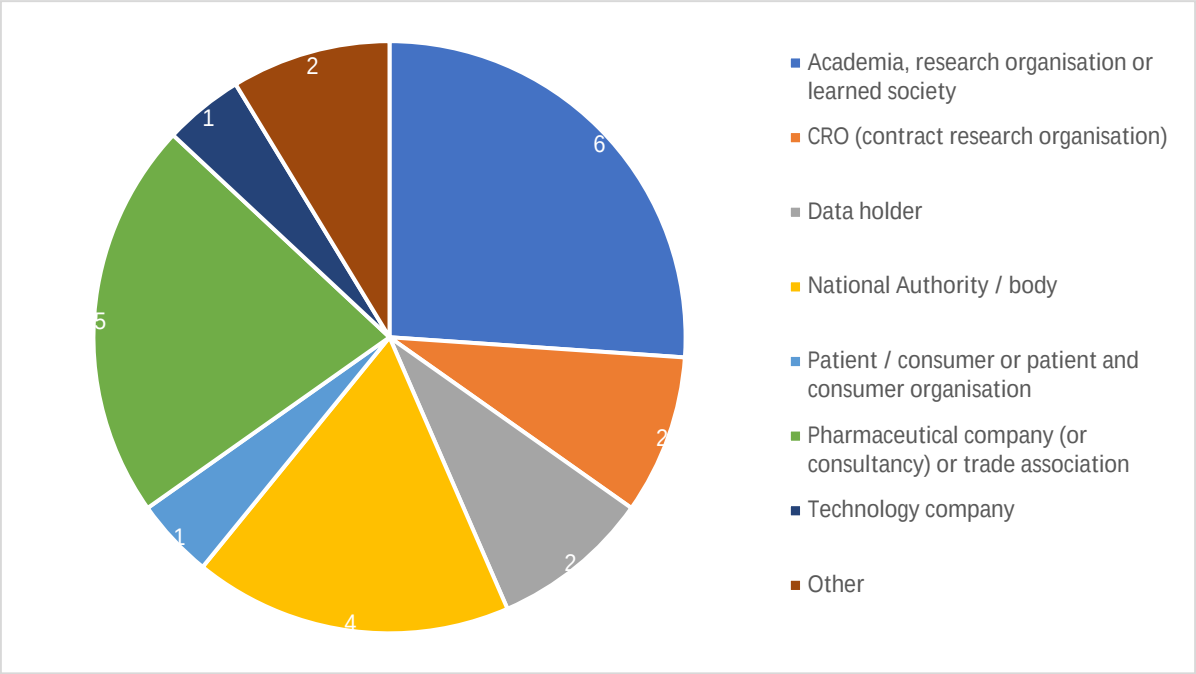
In the updated GPG, sections that included descriptions of the metadata (Section 5), registering a data source in the catalogue (Section 6) and maintenance of information in the catalogues (Section 7) have been removed, as suggested in some comments. These sections have been further elaborated and published in a separate document named 'User guide of the HMA-EMA Catalogues of real-world data sources and studies', hereafter referred to as the 'User guide'. The User guide provides descriptions of the catalogues' data fields, the validation process of the records submitted, as well as guidance on how to submit and maintain records in the catalogues.

The HMA-EMA Catalogues of real-world data sources and studies was launched in February 2024 and several suggestions received during the GPG consultation were incorporated in the first release of the system. Some comments are also addressed in the Support page of the catalogues.

2. Contributors

In total, 23 contributions were received from a wide range of stakeholders. The distribution of respondents by stakeholder type is reflected in the chart below.

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 An agency of the European Union



3. Summary of the main points raised during the consultation

The public consultation on the 'Good Practice Guide for the use of the HMA-EMA Catalogues of real-world data' (GPG) yielded a variety of comments and suggestions from stakeholders. All feedback received during the consultation has been analysed. Overall, respondents expressed support for the initiative and provided valuable input on how to enhance the usefulness and clarity of the guide. The main themes identified in the responses include aspects related to data quality, metadata elements, the scope of the catalogues, usability, the registration and maintenance of data sources as well as clarity and editorial improvements. Comments pointing out sections that would benefit from clarifications, further examples or better phrasing have been incorporated into the text accordingly. Due to the volume and diversity of the comments, this summary report will focus on the recurrent topics or key issues raised by stakeholders.

3.1. Data quality aspects

Several comments highlighted the need to align the concepts, terminology and definitions related to data quality mentioned in the GPG with the Data quality framework for European Union medicines regulation (EMRN DQF). Section 4 of the updated GPG has been revised accordingly.

3.2. Metadata elements, look-up values and vocabularies used in the catalogues

A significant number of comments focused on types of metadata elements that should be collected in the catalogues as well as updates to the look-up values and vocabularies used in the catalogue of data sources. These comments have mainly been addressed in the User guide and the List of metadata for the HMA-EMA Catalogues of real-world data sources and studies.

- Respondents suggested additional metadata elements to be collected to support the use cases outlined in the GPG and to more comprehensively describe a wide variety of data sources. Many of the suggestions will be taken into account in future updates of the List of metadata for

the HMA-EMA Catalogues of real-world data sources and studies and the User guide. Examples include:

  - Metadata elements indicating whether dates of diagnosis, prescription, dispensing, medicines administration/use are available.
  - Metadata related to the data collection process, actors involved and the data audit trail.
  - The date when various types of information started to be collected in the data sources, particularly as some sources expand the data they collect over time.
  - Capture whether "synthetic" versions of the dataset are available or can be created.
  - Elements created by the data holder, such as proprietary algorithms (e.g. disease progression for oncology patients).
  - Metadata elements to capture whether the data sources collect information on 'Specific medication/Treatments', which would be useful for patient registries where enrolment is based on exposure/treatment with specific medicines.

More detailed metadata or additional stratification variables related to cause of death, pregnancy and neonates, lifestyle factors, medical devices, sociodemographic information, education, informed consent to use the data for research and patient-generated data.

  - The method used by the data source to establish familial linkage.
  - The primary aim of data collection for the data source.
  - More free-text fields related to various metadata elements to allow the data holder to specify further details.
  - Metadata indicating whether data about diagnostic images are collected in the data source.

- Updates to the descriptions of metadata elements, including clarifications of terms have been implemented. For example, updates include how to register data source name aliases and the explanation of terms such as 'active population', 'data source' and 'underlying population'. Other comments received concerned the addition of more descriptions for values describing sociodemographic information collected in the data source (e.g. ethnicity, type of residence, rural area), as well as the metadata elements 'healthcare provider' and 'units of healthcare utilisation'. These suggestions will be considered in future releases of the catalogues

- Several respondents recommended using additional vocabularies alongside MedDRA for codifying diseases in the catalogues. This metadata element should be populated when the data source focuses on specific diseases (e.g. a patient registry). In these cases, the relevant disease(s) should be specified selecting few values from the MedDRA look-up. If the data source does not focus on specific diseases, a mapping to MedDRA is not necessary. MedDRA is the code list used in the EMA SPOR system to disease codification.

- Updates to the look-up values for various metadata elements, such as procedures, indications, diagnoses and medical events and cause of death have been implemented and several values have been added (e.g. ICD-11, SNOMED CT, OPS, OPCS, OMIM). The genetic data vocabulary has also been updated with additional values (e.g. HGNC and HGVS), as well as quality of life measurements (e.g. SF-12, SF-36). Further suggestions for updating the look-ups will considered in future releases. For example, adding values for 'data source type' or 'care setting' (e.g. for sources focusing on educational outcomes, child development and social care).

- It was highlighted that access to raw data and computational resources may be needed for a more in-depth assessment of intrinsic aspects of data quality. This could include verifying records and values, checking data against reference or plausible values and conducting other

computations. Such assessments could be performed by the data holders and periodically updated, preferably using automated tools. Data holders are encouraged to make the methods and the results of these assessments publicly available to support the assessment and replication of studies.

3.3. Scope of the catalogues

Several comments were raised regarding to the scope of the catalogues of data sources, particularly:

- Clarification on whether non-EU based data sources can be registered in the RWD Catalogue of data sources. This is clarified in the Support page of the catalogues (see section: FAQs on data source records).
- A clearer distinction between primary and secondary data collection, with clarifications on their definitions and their relevance to the scope of the RWD Catalogues. The scope of the catalogues has been updated in section 3.1 of the GPG.

Further information on the data sources within the scope of the catalogues can be found in the User guide (section 3: Registering a data source).

3.4. Registration, maintenance and usability of the catalogues

The GPG, as circulated for public consultation, included limited details on the maintenance of the catalogues as both the business processes and the catalogues were still under development. Several comments raised important topics regarding the maintenance of the catalogues, emphasising its significance for their quality and utility. In particular, respondents requested further details on how data holders will be incentivised to provide and periodically update metadata, what happens to data records if not updated within a certain period, the type of quality checks performed to the records submitted in the catalogues and how the evolution of metadata will be managed. The importance of data holders being responsible for registering data sources in the catalogue to ensure high-quality records was highlighted.

Several questions were raised concerning the registration process and the maintenance of the catalogues. Information regarding registration and maintenance processes has been included in the Support page of the catalogues and the User guide.

Clarifications were also requested regarding data protection. All information related to data protection can be found in the Privacy policy page of the catalogues.

Key aspects regarding the value and usability of the catalogues were also highlighted, such as the importance of comprehensiveness of data records, clear presentation of relevant and harmonised information “at a glance”, the systematic availability of comparable and accurate data for each data source (including the ability to compare data sources ‘side-by-side), the user friendliness of the system, search functionality, availability of dashboards and interoperability with other catalogues. These aspects are being considered in the ongoing improvement of the catalogues.

4. Next steps

The input provided by consultation participants has significantly shaped the content and structure of the GPG. The updated GPG has been agreed by the Methodological Working Party (MWP) and adopted by the Committee for Medicinal Products for Human Use (CHMP). Many of the suggestions related to the usability of the catalogues and metadata list will be considered in future releases of the catalogues

and future versions of the User guide, ensuring that the system and guidance continues to evolve in line with the needs of stakeholders.