

## 101191967 - GREG

### Testing, improving, and co-creating Guidance and Tools for Real World Evidence Generation and Use for Decision-Making in Europe

#### WP6 – RWD and Federated Analytics

## 6.2 Data Landscaping Reports

|                     |                                                                            |
|---------------------|----------------------------------------------------------------------------|
| Lead contributor    | Lucía Bellas-Fernández (8. EDHEN-F)                                        |
| Other contributors  | Lucía A. Carrasco-Ribelles (8. EDHEN-F)<br>Daniel Prieto-Alhambra (I. EMC) |
| Due date:           | 30 April 2026                                                              |
| Delivery date:      | 30 April 2026                                                              |
| Dissemination level | Public                                                                     |

This project is supported by the Innovative Health Initiative Joint Undertaking (IHI JU) under grant agreement No 101191967. The JU receives support from the European Union's Horizon Europe research and innovation programme and COCIR, EFPIA, Europa Bio, MedTech Europe, and Vaccines Europe.

|                                                                                   |                                          |                      |      |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |      |
|                                                                                   | <b>WP6</b> - RWD and Federated Analytics | <b>Version:</b> v4.0 |      |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | 2/20 |

# Contents

|                                                     |    |
|-----------------------------------------------------|----|
| Contents .....                                      | 2  |
| Document History .....                              | 3  |
| Definitions .....                                   | 4  |
| List of abbreviations .....                         | 6  |
| Executive Summary.....                              | 7  |
| 1 Introduction.....                                 | 8  |
| 2 Methods .....                                     | 9  |
| 2.1 Definition of data requirements .....           | 9  |
| 2.2 Identification of potential data partners ..... | 9  |
| 2.3 Landscaping .....                               | 9  |
| 2.4 Data partner metadata review .....              | 10 |
| 3 Results .....                                     | 10 |
| 4 Discussion and conclusion .....                   | 13 |
| Annexes.....                                        | 15 |
| Annex I Exploration phase reporting template.....   | 15 |

|                                                                                   |                                          |                      |      |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |      |
|                                                                                   | <b>WP6</b> - RWD and Federated Analytics | <b>Version:</b> v4.0 |      |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | 3/20 |

## Document History

| Version | Date       | Description                                                                 |
|---------|------------|-----------------------------------------------------------------------------|
| V1.0    | 09/03/2026 | First draft internal review.                                                |
| V2.0    | 18/03/2026 | Version for formal review by Katia Verhamme and Francisco Estupiñán-Romero. |
| V3.0    | 16/04/2026 | Version for formal review by GA.                                            |
| V4.0    | 30/04/2026 | Final version.                                                              |

|                                                                                   |                                          |                      |             |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|-------------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |             |
|                                                                                   | <b>WP6 - RWD and Federated Analytics</b> | <b>Version: v4.0</b> |             |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | <b>4/20</b> |

## Definitions

Participants of the GREG Consortium are referred to herein according to the following codes:

1. Erasmus Universitair Medisch Centrum Rotterdam (EMC)
2. The Chancellor, Masters and Scholars of The University of Oxford (UOXF)
3. SYNAPSE Research Management Partners SL (SYNAPSE)
4. National Institute for Health and Care Excellence (NICE)
5. GETREAL Institute (GetReal)
6. Societe Europeenne De Cardiologie (ESC)
  - 6.1. Region Uppsala (REGION UPPSALA)
7. Erasmus Universiteit Rotterdam (EUR)
8. Stichting European Health Data and Evidence Network (EHDEN-F)
9. Stichting Eupati Foundation (EUPATI)
10. Instituto Aragonés De Ciencias De La Salud (IACS)
11. University Of Galway (GALWAY)
12. The Hyve BV (THE HYVE)
13. University Of Dundee (UNIVDUN)
14. Universidade De Aveiro (UAVR)
15. Direktoratet for medisinske Produkter (NOMA)
16. Novo Nordisk A/S (NOVO)
17. Sanofi-Aventis Recherche & Developpement (SARD)
18. Bayer Aktiengesellschaft (BAYER)
19. Edwards Lifesciences Belgium BV (EW)
20. Medtronic Bakken Research Center B.V. (MDT)
21. Mölnlycke Health Care AB (MHC)
22. Institut De Recherches Internationales Servier (IRIS)
23. Pfizer Inc (PFIZER INC)
24. W. L. Gore & Associati S.R.L. (GORE S.R.L.)
  - 24.1. W. L. Gore et Associés S.A.R.L. (Gore S.A.R.L.)
  - 24.2. W. L. Gore & Associates BV (GORE BV)
25. Bristol-Myers Squibb Company Corp (BMS)
26. Boehringer Ingelheim International GmbH (BII GMBH)
27. Janssen Cilag SA (JCSPAIN)
28. GlaxoSmithKline Research & Development Limited (GSK)
29. AMGEN (AMGEN)
30. Sanofi US Services Inc. (Sanofi US Inc)
31. Sanofi-Aventis gulf F.Z.E. (SANOFI Gulf)
32. Sanofi Winthrop Industrie (SWI)
33. Sanofi Pasteur SA (Sanofi P SA)
34. Aventis Pharma Limited (Sanofi Pharma)
35. Sanofi Pasteur Limited Canada (Sanofi P L)
36. Edwards Lifesciences Sarl (EW SARL)
37. Edwards Lifesciences LLC (EW LLC)
38. Pfizer Hellas S.A. (Pfizer Hellas)
39. Pfizer AB (Pfizer AB)
40. Janssen Research & Development, LLC (JRDLLC)

|                                                                                   |                                          |                      |             |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|-------------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |             |
|                                                                                   | <b>WP6</b> - RWD and Federated Analytics | <b>Version:</b> v4.0 |             |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | <b>5/20</b> |

41. Medical Device Business Services Inc (MDBSI)
42. Edwards Lifesciences SRL (EW SRL)
43. Edwards Lifesciences SL (EW SL)
44. Edwards Lifesciences SAS (EW SAS)
45. Edwards Lifesciences GmbH (EW GMBH)
46. Medtronic, Inc. (MDT Inc)
47. Medtronic France (MDT FR)
48. Medtronic Italia S.p.A. (MDT IT)
49. Medtronic Ltd (MDT LTD)
50. Pfizer R&D UK Limited (PFIZER R&D UK)
51. Pfizer Limited (PFIZER LTD)
52. Actelion Pharmaceuticals Ltd (ACTELION)
53. Medical Devices & Diagnostics Global Services, LLC (MDDGB)
54. The University of Edinburgh (UEDIN)

|                                  |                                                                                                                                                                                                                            |
|----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Grant Agreement (GA)</b>      | The agreement signed between the beneficiaries and the IHI for the undertaking of the GREG project (101191967)                                                                                                             |
| <b>Project</b>                   | The sum of all activities carried out in the framework of the Grant Agreement.                                                                                                                                             |
| <b>Consortium</b>                | The GREG consortium, comprising the above-mentioned legal entities hereby referred as Participants of the GREG consortium.                                                                                                 |
| <b>Consortium Agreement (CA)</b> | Agreement concluded amongst GREG beneficiaries for the implementation of the Grant Agreement (GA). Such an agreement shall not affect the parties' obligations to the Community and/or to one another arising from the GA. |

|                                                                                   |                                          |                      |             |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|-------------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |             |
|                                                                                   | <b>WP6</b> - RWD and Federated Analytics | <b>Version:</b> v4.0 |             |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | <b>6/20</b> |

## List of abbreviations

| Abbreviation | Explanation      |
|--------------|------------------|
| HF           | Heart failure    |
| DP           | Data Partner     |
| UC           | Use Case         |
| EHDEN-F      | EHDEN-Foundation |

|                                                                                   |                                          |                      |             |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|-------------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |             |
|                                                                                   | <b>WP6 - RWD and Federated Analytics</b> | <b>Version: v4.0</b> |             |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | <b>7/20</b> |

## Executive Summary

This deliverable describes the first phase of a feasibility assessment, i.e., exploration phase, used in the IHI-GREG project to identify potential data partners (DPs) for real-world evidence studies within the IHI-GREG projects' consortium and EHDEN Network. This phase includes a preliminary data assessment phase, data landscaping phase and generation of the “Exploration report” after holding an “Exploration meeting”. This is prior to any institutional review board approval and can be developed in parallel to the study protocol.

In this document we will provide a whole overview of the process and an example of all the phases of a specific use case. The exploration phase focuses on assessing the study's needs in terms of data and the availability of relevant data sources before initiating a detailed feasibility assessment, which would involve DP contracting. The exploration results allow the identification and prioritisation of DPs within the IHI-GREG consortium and EHDEN Network. that can be invited to participate in subsequent feasibility or study phases.

The process involves screening data partners according to predefined study requirements and performing a landscaping exercise to evaluate whether relevant clinical concepts are present across the data sources that meet these requirements. The exploration phase (first phase of the feasibility assessment) is a broad, inclusive screening step, with conservative exclusion criteria, followed by a more detailed assessment to support final selection decisions. The workflow includes three main steps: identification of data requirements, systematic filtering of eligible data sources based on available metadata, and execution of landscaping queries to estimate the availability of relevant concepts.

This deliverable describes the exploration methodology and illustrates the process using a heart failure use case as an example. The exploration results allow the identification and prioritisation of data partners that can be invited to participate in subsequent feasibility or study phases.

|                                                                                   |                                          |                      |             |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|-------------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |             |
|                                                                                   | <b>WP6</b> - RWD and Federated Analytics | <b>Version:</b> v4.0 |             |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | <b>8/20</b> |

## I Introduction

Within the GREG project, use cases are evaluated using a structured workflow that supports the identification of suitable data sources. In the GREG project, the feasibility of use cases is evaluated in different domains, one of which is ‘data’.

Data feasibility assesses whether suitable data sources exist to answer the research question framed using the PICO structure (Population, Intervention/Exposure, Comparator, Outcome). This involves verifying both the availability of relevant data sources and whether they contain the information required to perform the statistical analyses needed to address the research question.

The consortium is jointly developing a standardized feasibility assessment framework with input from academia, industry, regulators, and HTA bodies to ensure it works across Europe’s varied real-world evidence settings. The framework is being built through an iterative cycle: creating a prototype, piloting it across medicines, devices, and combination products, gathering stakeholder feedback, and refining it repeatedly. A continuous learning system supports this process, using performance indicators, internal peer review, version control, and a shared lessons-learned repository for Regulatory and HTA audiences.

The first step in the data-feasibility procedure is the exploration phase—which includes (1) the preliminary data assessment, (2) the data landscaping, and (3) the generation of the Exploration Report. The purpose of this phase is to identify and shortlist candidate DPs that are potentially capable of supporting the study.

To carry out the shortlisting process, the data information requirements are first identified. This information must be translated into a data model specification (in this case, OMOP Common Data Model (CDM)). Secondly, data sources that meet these data information requirements are located, ideally using structured catalogues—such as the GREG catalogue or the EHDEN portal—to ensure that relevant and well-documented sources are systematically identified. The metadata contained in these catalogues is continuously fed and updated by DPs, ensuring that the information they contain remains current and reliable. Next, the exploration phase continues with clinical concept-based queries that operate on the previously collected metadata to finalise the identification of potential data partners. These concept-based queries are formulated using OMOP concept identifiers, ensuring consistent and semantically aligned retrieval across datasets. In order to optimise resources, the exploration phase is conducted prior to any contracting of DPs.

The need to define a standard process for shortlisting data partners arises from the growing complexity and scale of multi-database RWE studies. As the number of such studies increases, there is a need for approaches that are systematic, reproducible, and transparent to support data source selection. As explained below, this process reflects all the steps and decisions involved in DP shortlisting, documenting the criteria (i.e., lack of data or prioritisation of other use cases) behind these decisions.

This deliverable describes the exploration process, which remains the first part of a feasibility assessment, used in GREG and illustrates its application using a heart failure use case as an example.



|                                                                                   |                                          |                      |             |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|-------------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |             |
|                                                                                   | <b>WP6 - RWD and Federated Analytics</b> | <b>Version: v4.0</b> |             |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | <b>9/20</b> |

## 2 Methods

The exploration phase follows a structured workflow designed to systematically and reproducibly identify suitable data partners. A template has also been developed for reporting the steps mentioned below in a systematic way, which can be found in Annex I.

### 2.1 Definition of data requirements

The exploration process begins by extracting the research question and other key information data requirements from the use case description in the study outline. These requirements include:

- healthcare setting (i.e., data sources including healthcare information from primary care setting, hospital, claims)
- temporal coverage of the data (according to the proposed study timelines in the study outline) geographical scope (data from specific European countries: i.e. southern European countries)
- availability of specific data domains (meaning OMOP clinical domains: high-level classification of clinical data, such as condition, drug, procedures, observations, etc)

These requirements define the criteria used to systematically screen potential DPs within the data partner network.

### 2.2 Identification of potential data partners

DPs are screened against the predefined information data requirements using available information on the GREG catalogue or EHDEN portal. DPs that do not meet all these criteria are excluded from further consideration in this process. This selection process is documented through a flowchart to ensure transparency and reproducibility.

The output of this step is a flowchart reporting how many data sources meet each data requirement (from Section 2.1) and the list of data sources that meet all the data requirements for the use case.

### 2.3 Landscaping

For data partners meeting the data requirements, a landscaping exercise is performed to evaluate whether clinical concepts relevant for the specific use case are present in the data sources. These concept identifiers are provided by the use case lead(s) and are usually related to the study population, exposures, outcomes, or procedures (including potential confounders).

Landscaping is performed by querying a database (either the GREG catalogue or the EHDEN portal) that contains all patient counts for all possible concepts that were collected from each data partner during onboarding. The results provide an overview of the presence and approximate size of relevant patient populations across data partners.

Both the identification of potential data partners (section 2.1) and the landscaping (2.3) are conducted programmatically using the EHDEN Foundation's Landscaping R package. The Landscaping package code is version-controlled, and for producing each exploration report, the version of the package used is tracked, as is the version of the database being queried. The code for producing the content of the report is also version-controlled and available for the Consortium members.

|                                                                                   |                                          |                      |       |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|-------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |       |
|                                                                                   | <b>WP6</b> - RWD and Federated Analytics | <b>Version:</b> v4.0 |       |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | 10/20 |

## 2.4 Data partner metadata review

In addition to concept-based counts, additional metadata are reviewed to support the prioritisation of data partners. These include:

- data quality metrics, as produced by the OHDSI Data Quality Dashboard (tool) informing on plausibility, conformance and completeness which are automated checks evaluating whether data are sensible, correctly structured, and sufficiently completeness
- contractual readiness indicating whether a DP has a current framework agreement with EHDEN or an existing contract in force enabling data access for research purposes
- expected timelines for participation in studies, meaning availability of resources and information on performance previously gathered

This information supports the selection of data partners that are both technically suitable and ready to participate.

## 3 Results

The exploration phase produces several outputs that support transparent and reproducible data quality assessment and DP selection. These include:

1. DP screening flowcharts: Flowcharts summarise the filtering process applied to the data partner catalogue and document how many databases were excluded at each step (Figure 1).
2. Landscaping tables: Tables report the aggregated patient rounded counts obtained for the selected concept identifiers across potential data partners. These counts provide an initial indication of whether relevant populations can be identified in each database.
3. DP summaries: Additional tables summarise metadata related to data quality, contractual readiness, and expected participation timelines

Together, these outputs provide the information needed to prioritise candidate data partner. As an illustrative example, it was applied to a heart failure use case. The following data requirements were provided within the use case description:

- healthcare setting: any
- temporal coverage: data available from 2021 onwards
- geographical scope: no restriction
- required domains: visits, conditions, drug exposure, and death records

|                                                                                   |                                          |                     |                      |
|-----------------------------------------------------------------------------------|------------------------------------------|---------------------|----------------------|
|  | <b>D6.2 - Data landscaping reports</b>   |                     |                      |
|                                                                                   | <b>WP6 - RWD and Federated Analytics</b> |                     | <b>Version: v4.0</b> |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU | <b>11/20</b>         |

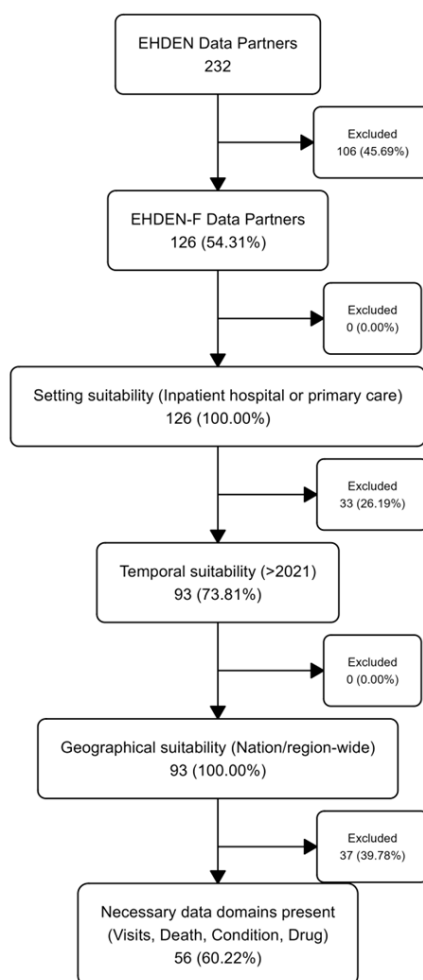


Figure 1. Flowchart of preliminary assessment of data requirements in EHDEN-F data partners

Using these criteria, the EHDEN data partner catalogue was screened to identify databases meeting the minimum requirements.

Landscaping was executed using concept identifiers provided by the UC lead related to heart failure diagnoses, cardiovascular conditions, and commonly used treatments. The resulting patient counts provided an overview of the presence of relevant populations across the eligible databases (Table 1).

Table 1. Patient counts obtained during Landscaping on the data partners meeting the preliminary assessment criteria (only selected rows)

|               | 2C<br>A-<br>Bra<br>ga | A<br>C<br>I | AKR<br>DB | AOU<br>PR | A<br>P<br>-<br>H<br>M | AP<br>DP | AR<br>CA | AZ<br>Ooste<br>nde | A<br>Z<br>K | BIF<br>AP | BIG<br>AN | Ba<br>rts |
|---------------|-----------------------|-------------|-----------|-----------|-----------------------|----------|----------|--------------------|-------------|-----------|-----------|-----------|
| Heart failure | 100                   | 23,300      | 3,300     | 1,500     | 19,300                | 0        | 7,400    | 3,700              | 2,100       | 643,900   | 87,200    | 33,300    |

|                                                                                   |                                          |  |                                  |
|-----------------------------------------------------------------------------------|------------------------------------------|--|----------------------------------|
|  | <b>D6.2 - Data landscaping reports</b>   |  |                                  |
|                                                                                   | <b>WP6 - RWD and Federated Analytics</b> |  | <b>Version: v4.0</b>             |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández |  | <b>Security:</b> PU <b>12/20</b> |

|                                           | 2C<br>A-<br>Bra<br>ga | A<br>C<br>I | AKR<br>DB | AOU<br>PR | A<br>P<br>-<br>H<br>M | AP<br>DP | AR<br>CA | AZ<br>Ooste<br>nde | A<br>Z<br>K | BIF<br>AP | BIG<br>AN | Ba<br>rts |
|-------------------------------------------|-----------------------|-------------|-----------|-----------|-----------------------|----------|----------|--------------------|-------------|-----------|-----------|-----------|
| Diastoli<br>c heart<br>failure            | 100                   | 0           | 0         | 100       | 0                     | 0        | 200      | 1,000              | 0           | 53,700    | 3,800     | 100       |
| Diastoli<br>c dysfunct<br>ion             | 100                   | 0           | 0         | 100       | 0                     | 0        | 200      | 1,000              | 0           | 57,100    | 4,100     | 100       |
| Left<br>heart<br>failure                  | 100                   | 5,000       | 1,200     | 1,500     | 19,300                | 0        | 7,400    | 700                | 300         | 118,000   | 6,100     | 33,300    |
| Systolic<br>heart<br>failure              | 100                   | 0           | 0         | 200       | 0                     | 0        | 500      | 200                | 0           | 45,000    | 5,500     | 100       |
| Acute<br>myocar<br>dial<br>infarctio<br>n | 100                   | 6,700       | 3,300     | 1,500     | 12,100                | 0        | 2,500    | 2,200              | 1,300       | 315,400   | 22,900    | 20,800    |
| Congest<br>ive heart<br>failure           | 100                   | 4,800       | 600       | 800       | 4,400                 | 0        | 2,200    | 1,200              | 900         | 338,900   | 32,000    | 13,800    |
| Chronic<br>heart<br>failure               | 100                   | 0           | 0         | 200       | 0                     | 0        | 900      | 1,000              | 0           | 53,700    | 5,500     | 200       |
| Perindo<br>pri usel                       | 200                   | 2,300       | 400       | 100       | 10,200                | 8,745    | 400      | 200                | 0           | 703,400   | 47,800    | 200       |

Finally, a table including summarised DP data quality, contractual readiness, and expected timelines was provided (Table 2). Currently, only the information on contractual readiness is available within the consortium.

Table 2. Data quality, contractual readiness, and expected timelines of the data partners meeting the preliminary assessment criteria.

| Data partner     | Data quality | Contractual readiness | Expected timelines |
|------------------|--------------|-----------------------|--------------------|
| <b>2CA-Braga</b> | NA           | FA                    | NA                 |
| ACI              | NA           | FA                    | NA                 |
| AKRDB            | NA           | FA                    | NA                 |
| AOU PR           | NA           | FA                    | NA                 |

|                                                                                   |                                          |                      |              |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|--------------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |              |
|                                                                                   | <b>WP6 - RWD and Federated Analytics</b> | <b>Version: v4.0</b> |              |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | <b>I3/20</b> |

| <b>Data partner</b> | <b>Data quality</b> | <b>Contractual readiness</b> | <b>Expected timelines</b> |
|---------------------|---------------------|------------------------------|---------------------------|
| AP-HM               | NA                  | FA                           | NA                        |
| ...                 | NA                  | MoU                          | NA                        |

*FA: Framework Agreement. MoU: Memorandum of Understanding. NA: Not Available.*

The exploration results allowed the identification of several candidate data partners that could potentially contribute to the study and informed the prioritisation of databases for the subsequent feasibility assessment. Having access to these results, during the “Exploration meeting” the following was decided for each data partner (go/ no go):

*Table 3. Comments for a go / no-go decision for all data partners meeting the preliminary assessment criteria during the “Exploration meeting”, which has not been held yet, will be provided in the following table.*

| <b>Data partner</b> | <b>Decision</b> |
|---------------------|-----------------|
| <b>2CA-Braga</b>    |                 |
| ACI                 |                 |
| AKRDB               |                 |
| AOUPR               |                 |
| AP-HM               |                 |
| ...                 |                 |

**Database acronyms:** Academic Clinical Center - Braga (2CA-Braga), Auria Clinical Informatics at Turku University (ACI), Akrivia Health (AKR DB), Azienda Ospedaliero-Universitaria di Parma (AO UPR), Health Data Warehouse of Assistance Publique - Hopitaux de Marseille (AP-HM), APDP Diabetes Portugal (AP DP), ARCA Cardiology Monasterio Foundation (AR CA), AZ Oostende - Data department (AZ OOstende), AZ Klina (AZK), BIFAP (Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (Pharmacoepidemiological Research Database for Public Health Systems)), Big Data in Healthcare, Aragon (BIGAN), Barts Health NHS Trust (Barts).

## 4 Discussion and conclusion

The exploration phase provides a structured approach to identifying potential candidate DPs for real-world evidence prior to conducting a feasibility assessment and the consequent study.

By combining metadata screening with concept-based landscaping, the process allows early assessment of whether relevant data sources exist and whether sufficient patient populations are available across databases. The exploration meeting will be recorded in order to nourish the language model implemented by other packages, in order to improve reproducibility.

The approach supports transparent and reproducible DP selection. The example of the heart failure use case illustrates how the exploration workflow can be applied in practice to identify potential databases for further feasibility assessment.

|                                                                                   |                                          |                      |              |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|--------------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |              |
|                                                                                   | <b>WP6</b> - RWD and Federated Analytics | <b>Version:</b> v4.0 |              |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | <b>I4/20</b> |

## References

Observational Health Data Sciences and Informatics. (n.d.). *Data Quality Dashboard*.  
<https://data.ohdsi.org/DataQualityDashboard/>

|                                                                                   |                                          |                      |       |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|-------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |       |
|                                                                                   | <b>WP6</b> - RWD and Federated Analytics | <b>Version:</b> v4.0 |       |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | 15/20 |

## Annexes

### Annex I Exploration phase reporting template

#### ASSESSMENT OF DATA FEASIBILITY

This section assesses the requirements on data (domain availability –conditions, drugs, measurements-, setting, geography, timelines, etc.), relevant feasibility counts, and reports the data partner selection process. Based on these, a recommendation on potential data partners is being made. The start date is reported for reproducibility purposes.

**Start date:** DD/MM/YYYY

#### Exploration report

This section reports the process for data partner shortlisting for the data feasibility assessment. From the submission form, the following information was extracted regarding the data requirements of this use case:

|                                           |                                                                                                    |
|-------------------------------------------|----------------------------------------------------------------------------------------------------|
| <u>Healthcare setting suitability</u>     | <i>To be completed with the UC setting needs, e.g., primary care</i>                               |
| <u>Temporal suitability</u>               | <i>To be completed with the UC temporal needs, e.g., data from &gt; 2021, data from &lt;2010</i>   |
| <u>Geographical suitability</u>           | <i>To be completed with the UC geographical needs, e.g., data from southern European countries</i> |
| <u>Presence of necessary data domains</u> | <i>To be completed with the UC data domains needs, e.g., conditions, drugs, and procedures.</i>    |

After exploring the EHDEN-F data partners specifications, the following flowchart has been produced reporting the number of data partners that do not meet each criterion. This selection process is **systematic** and transparent.

|                                                                                   |                                          |                      |              |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|--------------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |              |
|                                                                                   | <b>WP6</b> - RWD and Federated Analytics | <b>Version:</b> v4.0 |              |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | <b>I6/20</b> |

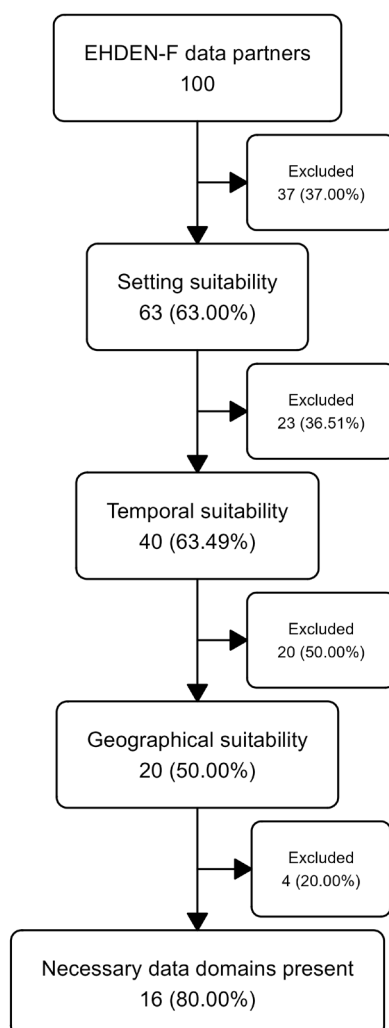


Figure 2. Flowchart of preliminary assessment of data requirements in EHDEN-F data partners – example only

Landscaping was carried out with the data partners that met all the requirements. As a result, a table reporting the patient counts for the selected concept IDs in these data partners is shown below.



|                                                                                   |                                          |                      |       |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|-------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |       |
|                                                                                   | <b>WP6 - RWD and Federated Analytics</b> | <b>Version:</b> v4.0 |       |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | 17/20 |

Table 4. Patient counts obtained during Landscaping on the data partners meeting the preliminary assessment criteria.

|            | DP1 | DP2 | DP3 | DP4 | DP5 | DP6 | DP7 | DP8 | DP9 | DP10 | ... |
|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|
| Concept 1  |     |     |     |     |     |     |     |     |     |      |     |
| Concept 2  |     |     |     |     |     |     |     |     |     |      |     |
| Concept 3  |     |     |     |     |     |     |     |     |     |      |     |
| Concept 4  |     |     |     |     |     |     |     |     |     |      |     |
| Concept 5  |     |     |     |     |     |     |     |     |     |      |     |
| Concept 6  |     |     |     |     |     |     |     |     |     |      |     |
| Concept 7  |     |     |     |     |     |     |     |     |     |      |     |
| Concept 8  |     |     |     |     |     |     |     |     |     |      |     |
| Concept 9  |     |     |     |     |     |     |     |     |     |      |     |
| Concept 10 |     |     |     |     |     |     |     |     |     |      |     |
| ...        |     |     |     |     |     |     |     |     |     |      |     |

|                                                                                   |                                          |                      |       |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|-------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |       |
|                                                                                   | <b>WP6</b> - RWD and Federated Analytics | <b>Version:</b> v4.0 |       |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | 18/20 |

The data quality, contractual readiness, and expected timelines of these same data partners are shown in the following table. Data quality is reported as a metric showing the percentage of tests passed, from an exhaustive list of data quality tests. Contractual readiness reports whether the data partner already has a memorandum of understanding or a framework agreement signed with the EHDEN-F. Timelines report the expected timelines for the data partner to participate in a study

*Table 5. Data quality, contractual readiness, and expected timelines of the data partners meeting the preliminary assessment criteria.*

| <b>Data partner</b> | <b>Data quality</b> | <b>Contractual readiness</b> | <b>Expected timelines</b> |
|---------------------|---------------------|------------------------------|---------------------------|
| DPI                 |                     |                              |                           |
| DP2                 |                     |                              |                           |
| DP3                 |                     |                              |                           |
| DP4                 |                     |                              |                           |
| DP5                 |                     |                              |                           |
| DP6                 |                     |                              |                           |
| DP7                 |                     |                              |                           |
| DP8                 |                     |                              |                           |
| DP9                 |                     |                              |                           |
| DPI0                |                     |                              |                           |
| ...                 |                     |                              |                           |

Having access to these results, during the “Exploration meeting” the following was decided for each data partner:

*Table 6. Comments on the data partners meeting the preliminary assessment criteria during the “Exploration meeting”. The decision will be a go / no go, based on the assessment previously carried before. The decision will be a consensum between the UC lead, WP4-5-6 leads.*

| <b>Data partner</b> | <b>Decision</b> |
|---------------------|-----------------|
| DPI                 |                 |
| DP2                 |                 |
| DP3                 |                 |

|                                                                                   |                                          |                      |       |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|-------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |       |
|                                                                                   | <b>WP6</b> - RWD and Federated Analytics | <b>Version:</b> v4.0 |       |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | 19/20 |

| <b>Data partner</b> | <b>Decision</b> |
|---------------------|-----------------|
| DP4                 |                 |
| DP5                 |                 |
| DP6                 |                 |
| DP7                 |                 |
| DP8                 |                 |
| DP9                 |                 |
| DP10                |                 |
| ...                 |                 |

Therefore, the list of data partners, sorted by preference, to be invited for the data feasibility assessment is as follows. A number of them were invited according to resource requirements.

1. Data partner to be invited
2. Data partner to be invited
3. Data partner to be invited
4. Data partner to be invited
5. Data partner to be invited
6. Data partner to be invited
7. Data partner to be invited
8. Data partner to be invited
9. Data partner to be invited
10. Data partner to be invited

|                                                                                   |                                          |                      |       |
|-----------------------------------------------------------------------------------|------------------------------------------|----------------------|-------|
|  | <b>D6.2 - Data landscaping reports</b>   |                      |       |
|                                                                                   | <b>WP6</b> - RWD and Federated Analytics | <b>Version:</b> v4.0 |       |
|                                                                                   | <b>Author(s):</b> Lucía Bellas-Fernández | <b>Security:</b> PU  | 20/20 |

## Annex II: Full Data Feasibility Process Outline

