TABLE
OF CONTENTS

PREFACE

First and foremost, we are fortunate to be coordinating a very dedicated, enthusiastic, and expert consortium of twenty-two partners, eleven public/eleven industry, especially through a global pandemic and a year of historic importance.

The consortium is steadfastly supporting the European research community in facilitating the introduction of 21st century research tools within an open science network, at scale. The EHDEN project is committed to success within the pillars of infrastructure, research, and community-building.

At the core of EHDEN is not data, technology, or digital innovation per se, but collaboration. Key to success is the project consortium, but critical to EHDEN is enabling partnerships between those Data Partners who generate and use clinical data, those wishing to perform research on this data, and those who can enable this relationship. We believe that EHDEN can have a significant impact on patient care in Europe by enabling federated research on health data through standardisation of data and analytical pipelines to generate timely and reliable evidence.

We hope you enjoy reading this summary report of our year 2 progress and are very grateful you are on board for EHDEN's exciting and impactful journey.

Peter Rijnbeek
EHDEN Coordinator
Associate Professor Health Data Science
Erasmus Medical Center, The Netherlands

Nigel Hughes
EHDEN Project Leader
Scientific Director RWD/RWE
Janssen R&D, Belgium

INFRASTRUCTURE

- Establishment of the EHDEN architecture (EHDEN Database **Catalogue**, data characterisation **tools** and **dashboards**) and proposed to launch publicly in year three.
- Contributed to **ETL** and **mapping tools**, further developed the **ARACHNE** tool, integrated security framework and advanced multiple analytical pipelines.

EDUCATION & COMMUNITY

- EHDEN **Academy** has over 800 participants globally and 9 different courses.
- **Value propositions** were finalised for key stakeholders and we analysed **sustainability** learnings of past and current projects
- We actively **engaged** with sister BD4BO projects, IMI, the EU Commission and EMA on areas of mutual relevance.

OPEN CALLS

- Processes and tools for the **open calls** are being fine tuned following each call.
- 61 **Data Partners** across 16 countries.
- 26 **SMEs** selected, trained, and certified.
- Launched a Rapid Collaboration call for Data Partners in response to the **COVID-19** pandemic.

RESEARCH & OUTCOMES

- Methodological research on Sample size, Natural Language Processing and AI.
- We ran **Study-a-thons** on Rheumatoid Arthritis therapeutics and on COVID-19 with OHDSI.
- Protocols for Drug Utilisation, Drug Safety, Health Technology Assessment **Use Cases** have been developed and ran on prior OMOP mapped data.
- Completion of the evaluation of the **ICHOM** standards for inclusion in the OMOP CDM.

MAJOR ACHIEVEMENTS

We are proud of the project's progress thus far, and on behalf of the consortium, thank everyone who has collaborated with us on our journey – it has been remarkable!

Certainly, the historic year of 2020 and the ongoing global pandemic have reinforced the need to ensure the right data is in the right place to answer the right question, within the appropriate sociotechnical framework and governance.

We have met all major milestones on time, as anticipated. Currently, five out of 34 deliverables were legitimately postponed and will be submitted in the months to come.

An overview of the deliverables can be found on our **Zenodo** page.

2020 OPEN CALLS

SECOND SME CALL

(February 2020)

**35 APPLICATIONS
15 SELECTED**

COVID-19 DATA CALL

(April-May 2020)

**75 APPLICATIONS
25 SELECTED**

GENERAL DATA CALL

(September-October 2020)

**34 APPLICATIONS
27 SELECTED**

SME CATALOGUE

Following our two open calls for SMEs, we now have a total of 28 SMes in our SME Catalogue. Each of these SMEs was selected by our independent SME Certification Committee in one of our open calls, received training via the EHDEN Academy and was certified during a two-day certification meeting.

EHDEN continues to build and strengthen this community via monthly update calls, an SME forum where experiences and knowledge can be shared, etc. to further strengthen this collaboration.

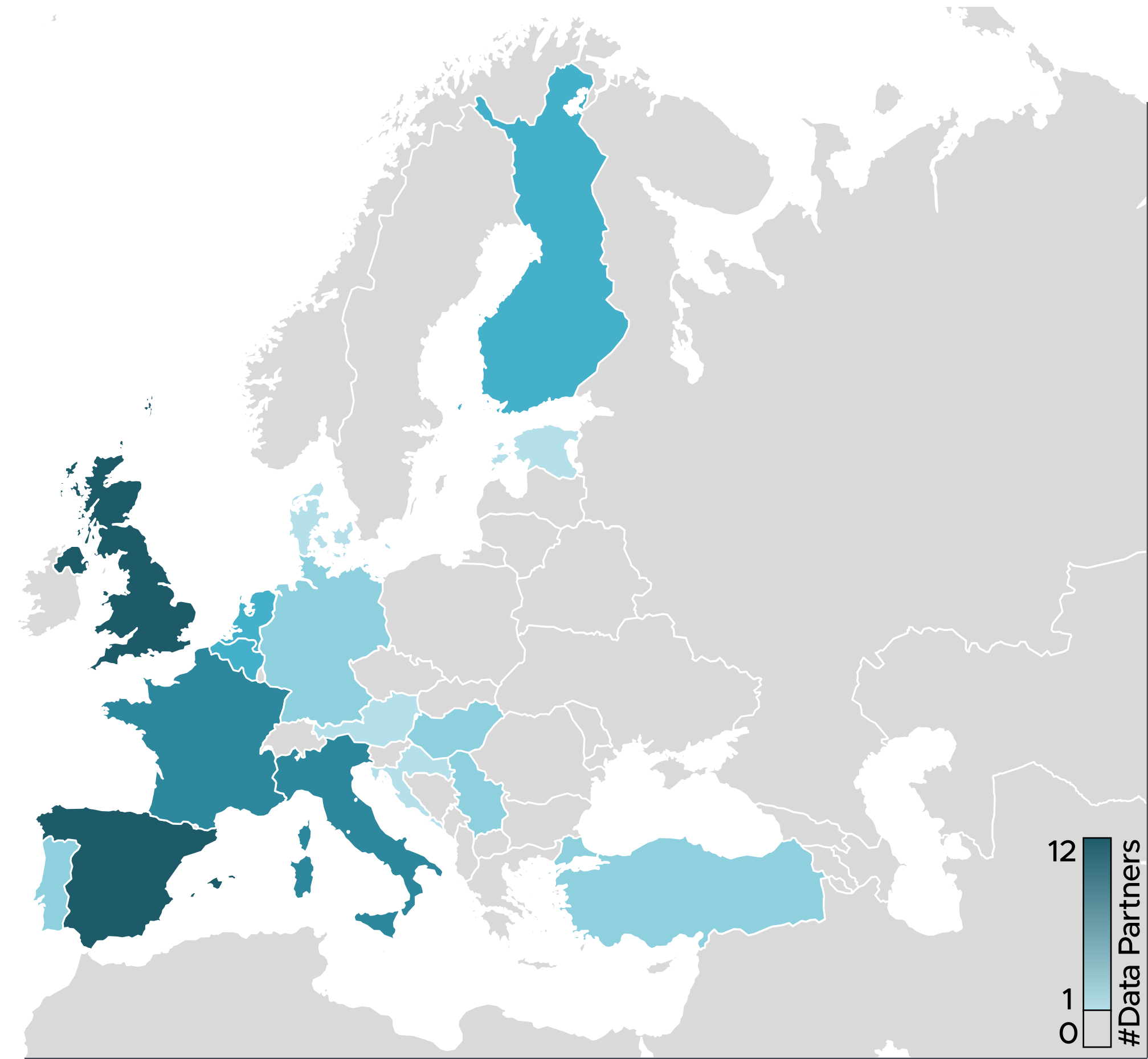Acces the catalogue via **ehden.eu/business-directory/**

FEDERATED NETWORK

The EHDEN project aims to collaborate with diverse institutions and data custodians across the EU, with a goal of harmonising source data to the OMOP common data model locally, within a federated network.

Following the 3 open calls to date we have organised, we currently have **61 data partners** from **16 different countries** which are mapping their data to the OMOP common data model. This includes a few EHDEN project partners who have also mapped their data to the OMOP CDM for use in the federated network.

The most up to date overview is availble at any time via **www.ehden.eu**

DATA PARTNERS

UNITED KINGDOM (12)

- CPRD
- DataLoch
- Health Informatics Centre
- King's College London
- Royal College of General Practitioners Research and Surveillance Centre
- SAIL Databank
- University College London
- Akrivia Health
- Barts Health NHS Trust
- Harvey Walsh Ltd
- Leeds Teaching Hospitals
- UK National Neonatal Research Database

FRANCE (7)

- APH de Marseille
- Bordeaux Hospital
- Bordeaux PharmacoEpi
- Cegedim Health Data
- CHU de Toulouse
- Health Data Hub
- Lille University Hospital

PORTUGAL (2)

- Rosa Maria Príncipe
- 2CA-Braga

CROATIA (1)

- Clinical Hospital Dubrava

AUSTRIA (1)

- Medical University of Vienna

SPAIN (11)

- Biocruces Bizkaia Health Research Institute
- Parc Salut Mar Barcelona
- Fundación de Investigación Biomédica del Hospital Universitario 12 de Octubre
- HULAFE
- FIIBAP
- Servicio Cántabro de Salud and IDIVAL
- The Information System for Reseach in Primary Care
- Vall d'Hebrón Hospital Marina Salud S.A.
- Parc Sanitari Sant Joan de Déu

ITALY (6)

- Fondazione IRCCS Istituto Neurologico Carlo Besta
- AUSL Reggio Emilia
- Fondazione Poliambulanza Istituto Ospedaliero
- Azienda Policlinico Universitaria di Modena
- Fondazione IRCCS Ca' Gran da Ospedale Maggiore Policlinico
- SIMG

SERBIA (2)

- Clinical Center of Serbia
- Clinical Centre of Nis

DENMARK (1)

- Center for surgical science

BELGIUM (4)

- Lynxcare
- Medaman
- AZ Klina
- University MS Center

THE NETHERLANDS (4)

- NICE foundation
- Amsterdam UMC
- IPCI
- STIZON

FINLAND (3)

- Clinical Biobank Tampere
- Hospital District of Southwest Finland
- BCB Medical Ltd

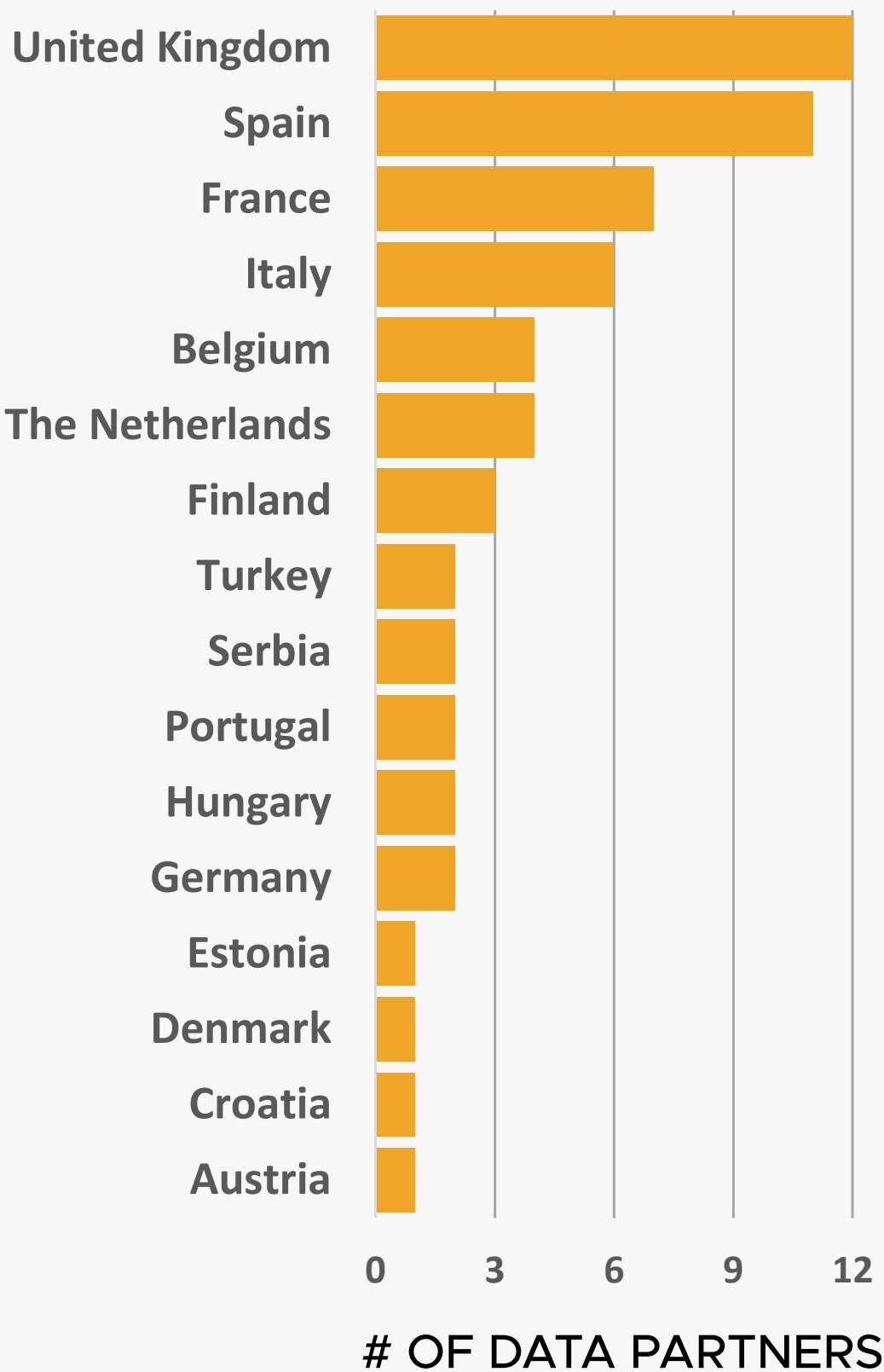## ESTONIA (1)

- University of Tartu

TURKEY (2)

- Istanbul University - Faculty of Medicine
- IUC Cerrahpaşa TIP Fakület-esi

GERMANY (2)

- Charité - Universitätsmedizin
- MS Forschungs- und Projek-tentwicklungs

HUNGARY (2)

- Semmelweis University
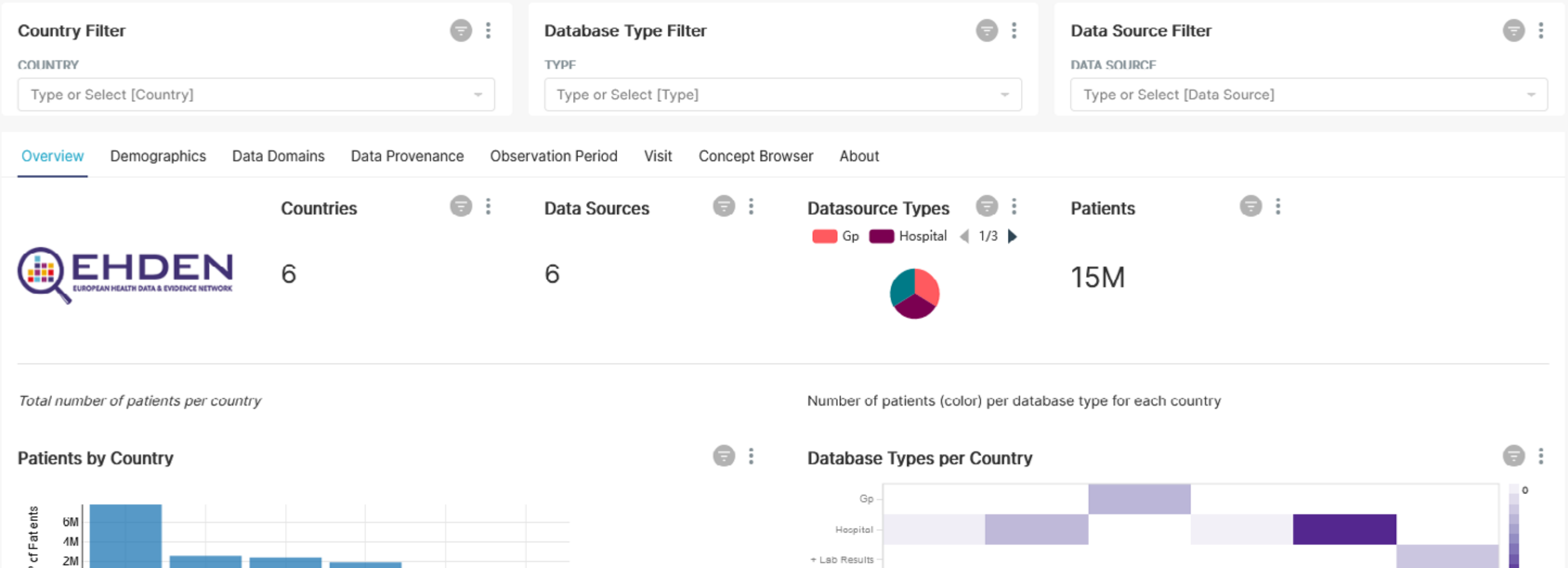- University of Pécs

OF DATA PARTNERS

7

EHDEN PORTAL

The EHDEN portal aims to enable the re-use of standardised health data across Europe and to support transparent and reproducible research. EHDEN envisions a socio-technical framework that enables researchers to find data, re-use data, and share analytical methods and study results.

The portal architecture is composed of multiple stand-alone applications, both existing open source OHDSI tools and EHDEN-developed tools, which we bring together. As such, the platform links to the Database Catalogue, the EHDEN Academy, ATLAS and Arachne and integrates the Elixir AAI (Elixir Authentication and Authorisation Infrastructure).

A full public release is anticipated in the second quarter of 2021 while we continue working on the portal functionallity and adding more data partners to the Database Catalogue.

The bottom image showsa part of one of the dashboards available via the EHDEN Portal.

TOOL AND METHOD DEVELOPMENT

A new **ARACHNE** version has been released, supporting fully customised R and SQL analysis. This has been implemented as we continue to work on the framework for federated studies.

New features were added to **White Rabbit** and **Rabbit in a Hat**. E.g., a new tool that performs an inspection on the CDM and automatically generates a report that contains quality measures.

An EHDEN **security policy** is being drafted on the basis of the ISO 27001 information security standard.

Developed a new **Data Network Dashboard** tool to aggregate information from distributed OMOP CDM databases using the CatalogueExport package to construct graphical dashboards.

Empirically determined **sample size** considerations for developing predictive models.

Implemented an R-package for assessing the **heterogeneity of treatment** effect, which predicts the optimal treatments at the individual level, accounting for an individual's risk for harm and benefit outcomes.
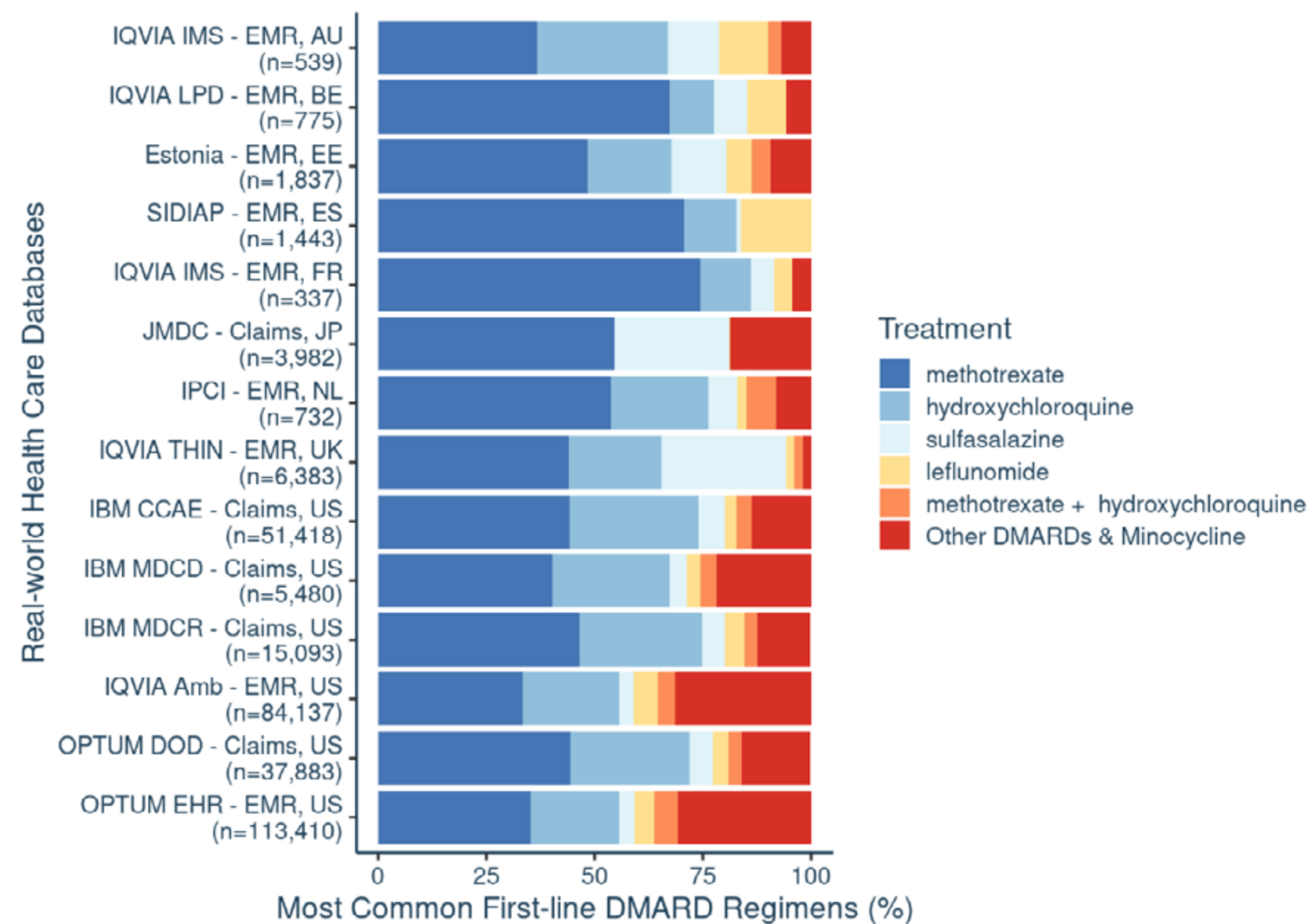
Published a publication on the role of explainability in creating trustworthy **artificial intelligence** for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies

Built a pipeline to assess the added value of **Natural Language Processing** of textual data in EHRs for improving patient-level prediction models.

(Left image) The use of first-line DMARDs in a total of 17 data sources, 11 countries, and including 7 European countries: Belgium, Estonia, France, Germany, the Netherlands, Spain, and the UK. This image depicts the existing great variability in the use of different DMARDs in first line after the diagnosis of RA in all the contributing data sources.

STUDY-A-THON
RHEUMATOID ARTHRITIS

Between January 13 and 17, 2020, approximately 40 people of all disciplines gathered in Barcelona for the second EHDEN Study-a-thon. This time, research questions were situated in the domain of Rheumatoid Arthritis as we evaluated drug utilisation of Disease Modifying Anti Rheumatic Drugs (**DMARDs**), their safety profiles and outcomes, as well as performing prediction modelling.
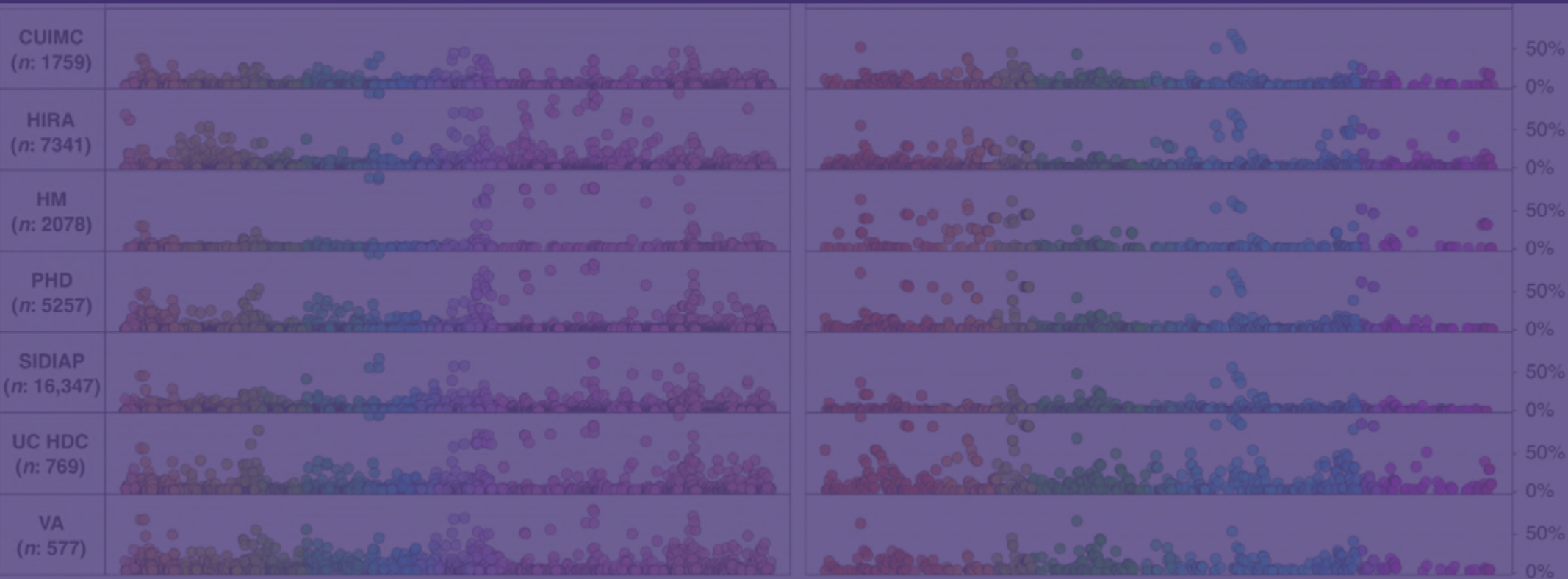
(**More info**)

STUDY-A-THON
COVID-19

In March 2020, as the world came to a halt due to the SARS-CoV-2 pandemic, clinicians, scientists, governments and the public all wanted to know more about the characteristics of patients with COVID-19, and if certain treatments were safe and effective. To this end, on 26-29 March, the Observational Health Data Sciences and Informatics (OHDSI) community, supported by EHDEN, hosted a COVID-19 virtual Study-a-thon wich was attended by more than 330 researchers, with thirty-seven healthcare databases from 30 different countries. In just 4 days, we managed to generate evidence which directly impacted COVID-19 care and patients.

PEER REVIEWED STUDIES

- **Hydroxychloroquine Safety Profile (Lancet Rheumatology)**
- **COVID19 Characterization Study (Nature Communications)**

PREPRINTS

- **COVID-19 Personalized Risk Calculator**
- **Renin-Angiotensin blockers and susceptibility to COVID-19**
- **Validation of the COVID-19 Vulnerability Index (C-19)**
- **Risk of depression, suicidal ideation, suicide and psychosis with hydroxychloroquine treatment for rheumatoid arthritis**
- **Characteristics and outcomes of COVID-19 patients with and without obesity**
- **Multinational COVID-19 Drug Utilization heterogeneity**
- **Characteristics, symptoms and outcomes of COVID-19 in pregnant woman**
- **Baseline phenotype and 30-day outcomes for COVID-19**
- **Effectiveness of Famotidine in COVID-19 patients**

MORE INFO

- Visit our **EHDEN Zenodo page** for the latest publications
- **OHDSI COVID-19 Study-a-thon page**
- EHDEN Website newspost **1**, **2** and **3**

USE CASE 1
DRUG UTILISATION

We continued the analyses related to the previously approved study of drug utilisation in respiratory diseases in previously mapped data including the Dutch IPCI primary care database and are ongoing in the UK CPRD GOLD data.

Preliminary analyses demonstrate the feasibility of using non-parametric, fully data-driven methods for the identification of changes in the population-level uptake of a given medicine over time. Both hypothesis (segmented linear regression) as well as data-driven methods (JointPoint regression and Bayesian change point detection) are available for the analysis of the impact of risk minimisation measures.

During the COVID-19 Study-a-thon, we initiated research on the utilisation of medicines to treat COVID-19. We performed both population-level and patient-level drug utilisation analyses.

USE CASE 2
DRUG AND DEVICE SAFETY

We have leveraged from existing tools previously created by the OHDSI community, including the Population Level Estimation (PLE) package, and added new tools, making them available to the wider community of EHDEN and beyond.

We organised, hosted, and completed a Study-a-thon on the comparative safety of RA treatment (Cf. Page 11).

In the early days of the COVID-19 pandemic, we completed the largest ever analysis of the cardiovascular safety of hydroxychloroquine in combination with azithromycin. This work has influenced regulatory and clinical decision making. Within weeks of this and following regulatory concerns of potential neuro-psychiatric safety issues with the use of hydroxychloroquine, we completed an additional analysis on the safety of hydroxychloroquine in terms of severe mental health disease.

USE CASE 3
HTA

The HTA use case on COPD was completed. We explored the current usefulness of the OMOP CDM and its associated analytical tools for generating evidence for use in HTA and provided recommendations for the future development of the model and tools (See our peer-reviewed publication on this topic).

We found that while the OMOP CDM does indeed support the efficient generation of evidence across otherwise disparate datasets, there were several important limitations that compromised the reliability of the generated estimates. For example, the source data often poses various challenges and the standard vocabularies and concepts that were used in the mapping did not appropriately reflect the nature of healthcare delivery in each country or support robust comparisons across countries. In addition, the analytical tools largely did not support the types of analysis required for HTA. These limitations are not inherent to the CDM approach, but rather reflect its current development and focus on regulatory use cases and the nature of the mapping processes implemented.
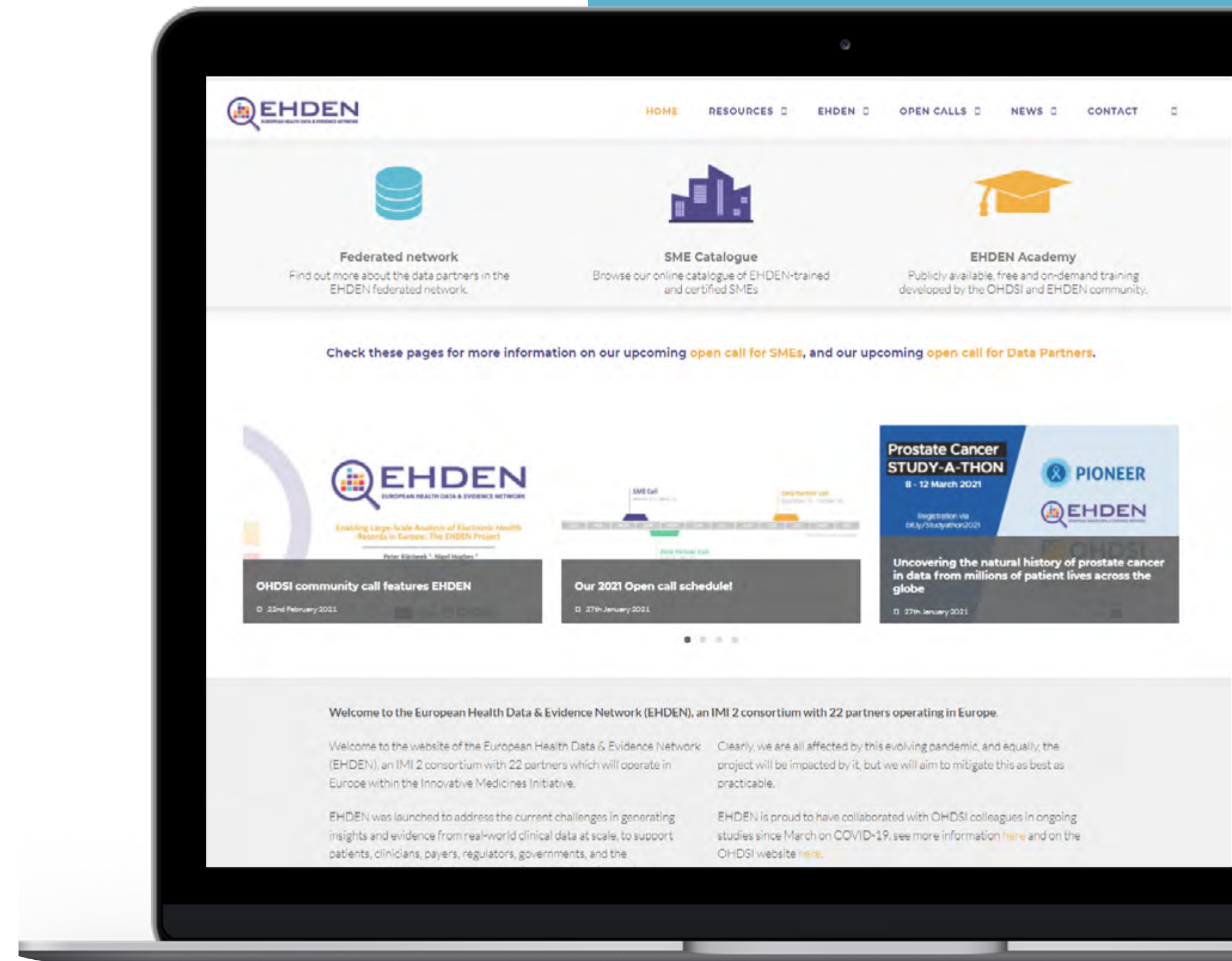
DISSEMINATION

EHDEN aims to be at the forefront of open science, making healthcare data more accessible for research. Our dissemination activities reflect this vision and as such, as of 2020, all our papers, preprints and public deliverables are now available via the **EHDEN Zenodo page.**

COMMUNICATION

As EHDEN is a socio-technical project, building a community that will endure and use the project outcomes long after the project has ended is one of the top goals of our communication strategy. This has resulted in an average of over 2000 visitors of the **EHDEN website** each month and a steady increase in followers on **Twitter** and **LinkedIn**. A complete overview of all newspost, newsletters, webinars, podcasts, movie materials, etc., can be accessed via our website.

EHDEN ACADEMY

Launched publicly in April 2020, the **EHDEN Academy** aims to become the trusted and go-to place for courses on using real-world data and generating real-world evidence.

There are currently 9 courses available in the Academy with more in development and planned throughout 2021. As the amount of courses further expands, so will the learning paths which we offer.

At the end of the second year, over 800 users from all over the world and with various professional backgrounds have started their learning journey in the Academy

EHDEN Foundation

Population-level Effect Estimation

Phenotype Definition, Characterisation and Evaluation

Extract, Transform and Load

OMOP CDM and Standardised Vocabularies

Patient-level Prediction

OHDSI-in-a-box

Infrastructure

ATLAS

COLLABORATIONS

BD4BO

- **HARMONY** and EHDEN coordinated their respective COVID-19 open calls for Data Partners.
- **PIONEER** and EHDEN have a co-working group to explore opportunities to collaborate around Data Partners, SMEs, research uses cases/study-a-thons, and educational support via the EHDEN Academy.
- We have MoUs in place with BD4BO sister projects, but also **HTx**, and have had numerous engagements with applicant consortium for IMI projects, as well as contributing to **Health Outcomes Observatory** and **NEURONET**.

EUROPEAN COMISSION & EMA

- We have been actively collaborating with the **European Commission** on the development of the European Health Data Space (EHDS) and the development of EHDEN.
- Discussions with **EMA** focused on the need for a European platform, largely adopted, which facilitates data analysis and RWE generation to support healthcare needs. The collaboration with the EMA explores the opportunities that are created by the OMOP CDM and OHDSI framework, collaborate on mapping activities for drug codes, inform them about methodological development, and to discuss convergent use cases of interest.
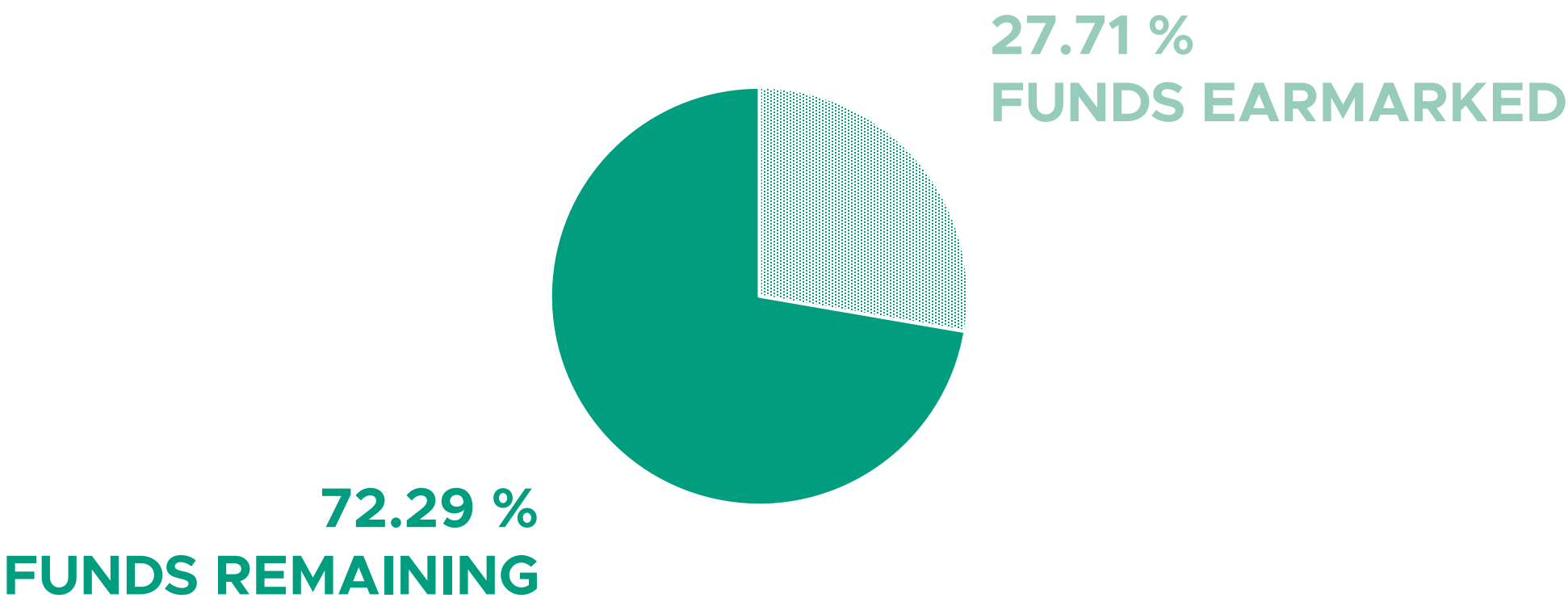
OHDSI

- We co-organised Study-a-thons, have jointly worked on methodological research, the data quality dashboard and lead or participated in the workgroups on Patient-Level Prediction and Population-Level Effect estimation.
- We collaborated with OHDSI on the development of analytical tools, and on the content of the EHDEN Academy, which is being promoted by OHDSI as the learning environment for the global community.
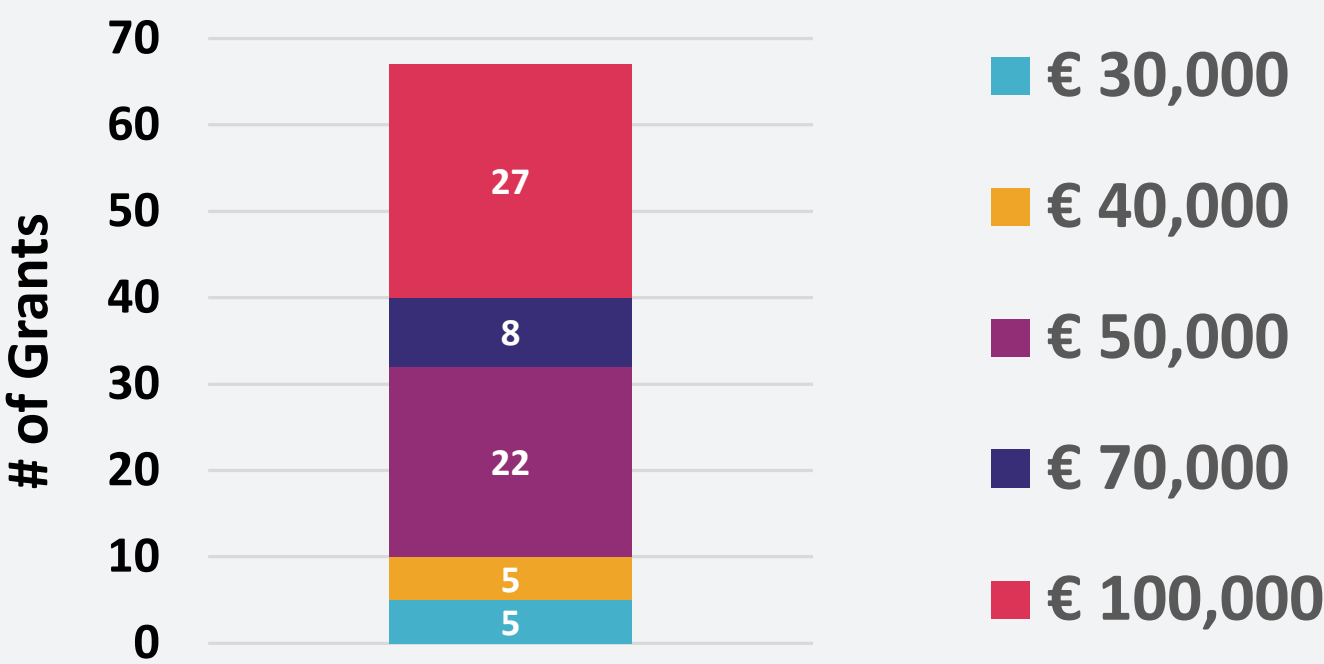
HARMONISATION FUND

The EHDEN Harmonisation fund has a total of 17 000 000 € available over the duration of the project. During the first year of the project, we held one open call for data partners for which 2 million euro was earmarked. In the second year of the project, we held two more open calls for data partners to which we allocated funding up to 1 million and 2 million euros respectively, equalling a total of 5 million euros from the three open calls to date.

(*Top graph*) Having held three open calls for data partners, we've granted 4 710 000 €, which means that we've currently earmarked 27.71% of the total funds available.

(*Bottom graph*) For the 3 open calls combined, 67 applications were eligible for financial support and the grant sizes, which depend on the size and complexity of the mapping, were distributed as indicated.

27.71 %
FUNDS EARMARKED

72.29 %
FUNDS REMAINING

- vv

Bottom chart (# of Grants):
- € 30,000: 5
- € 40,000: 5
- € 50,000: 22
- € 70,000: 8
- € 100,000: 27

SUSTAINABILITY

Building on the collective knowledge and experience in the consortium, we have been tackling the need for sustainability from the beginning. Following the initial discussions in the first year, a sustainability workshop was organised at the second general assembly meeting.

We performed a landscape exercise on sustainability efforts and successes of other IMI projects and to position EHDEN in the current RWE landscape. We developed Value Propositions for all identified EHDEN stakeholders and assessed how EFPIA and Academic partners value our list of potential value offerings (Certification authority, Tool development, EHDEN-in-a-box, Mapping services, Studies, Dashboarding/RWE journal, Honest broker, Scientific society, Community, Education & training).

All of these materials which we have developed are now being used as input to explore and develop the EHDEN sustainability model, which we hope to kick-off in the third year.

LOOKING FORWARD

We hope you will agree EHDEN has continued sterling work in year two, and we are looking forward to year three, particularly as we all hope to emerge from the COVID-19 pandemic.

We have some clear objectives now to meet our longer term goals, (1) to be able to collaborate with Data Partners on use cases following the successful completion of their mapping to the OMOP CDM, (2) to launch the EHDEN Portal, linked to the development of our 'research operating model', (3) to create our initial sustainability programme in parallel to the IMI project, and, (4) to collaborate wider with OHDSI on the uptake of the OMOP CDM and OHDSI framework, and community education in Europe.

We are indebted to our consortium partners and the incredible drive and enthusiasm exhibited in our collective goals in ensuring the right data is in the right place to answer the right questions, not just during a pandemic, but across healthcare. Our vision remains, enabling better health decisions, outcomes, and care.