HMA
Heads of Medicines Agencies

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Conceptual Framework

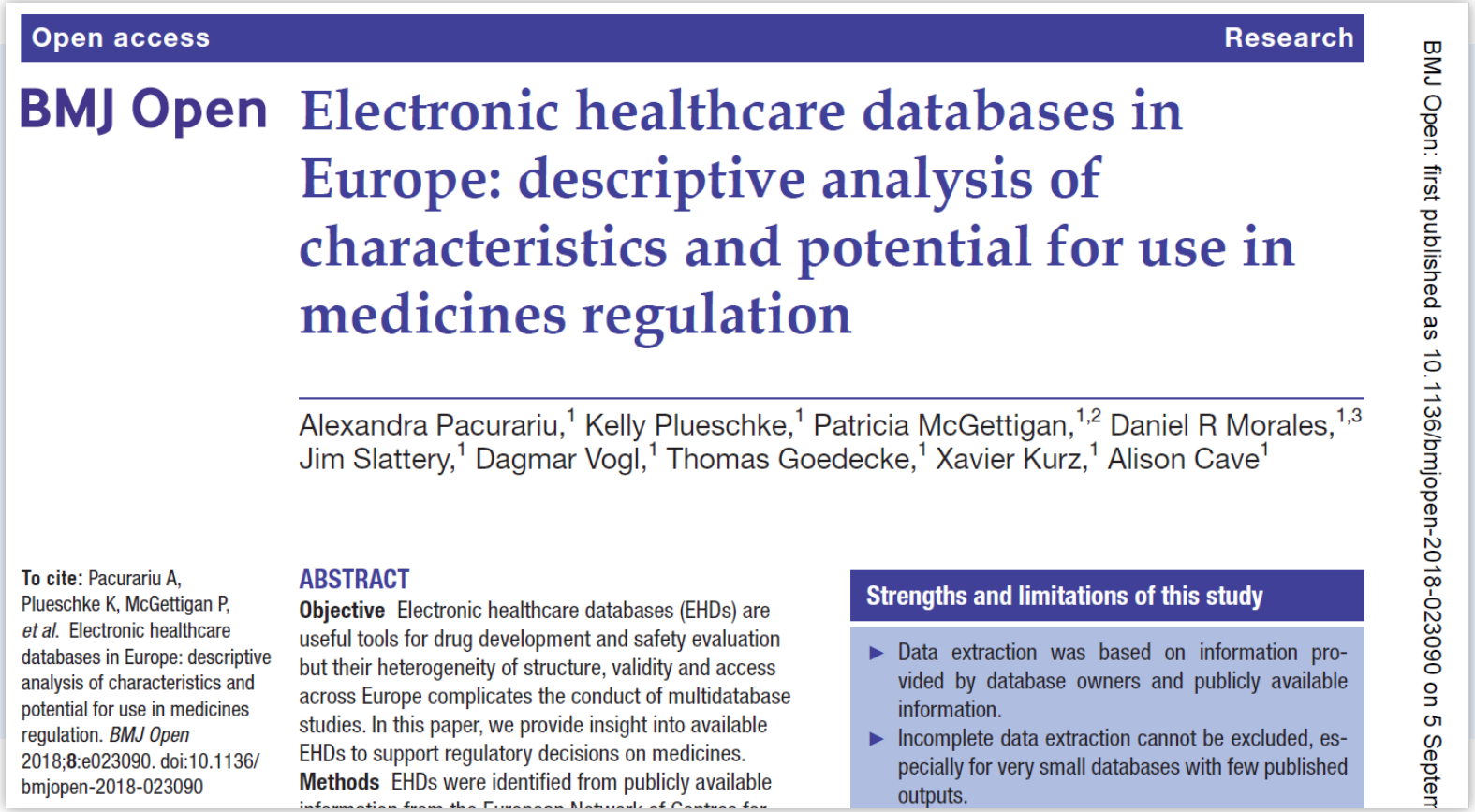
Technical workshop on real-world metadata for regulatory purposes
Virtual meeting, April 12, 2021

Presented by Dr. Rosa Gini
ARS Toscana

Contents

Contents

Conceptualising European data sources



Open access Research

BMJ Open

Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation

Alexandra Pacurariu,[1] Kelly Plueschke,[1] Patricia McGettigan,[1,2] Daniel R Morales,[1,3] Jim Slattery,[1] Dagmar Vogl,[1] Thomas Goedecke,[1] Xavier Kurz,[1] Alison Cave[1]

To cite: Pacurariu A, Plueschke K, McGettigan P, *et al*. Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation. *BMJ Open* 2018;**8**:e023090. doi:10.1136/bmjopen-2018-023090

ABSTRACT

Objective Electronic healthcare databases (EHDs) are useful tools for drug development and safety evaluation but their heterogeneity of structure, validity and access across Europe complicates the conduct of multidatabase studies. In this paper, we provide insight into available EHDs to support regulatory decisions on medicines.

Methods EHDs were identified from publicly available

Strengths and limitations of this study

- Data extraction was based on information provided by database owners and publicly available information.
- Incomplete data extraction cannot be excluded, especially for very small databases with few published outputs.

BMJ Open: first published as 10.1136/bmjopen-2018-023090 on 5 Septem

Conceptualising European data sources

Open access Research

BMJ Open

Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation

Alexandra Pacurariu,[1] Kelly Plueschke,[1] Patricia McGettigan,[1,2] Daniel R Morales,[1,3] Jim Slattery,[1] Dagmar Vogl,[1] Thomas Goedecke,[1] Xavier Kurz,[1] Alison Cave[1]

To cite: Pacurariu A, Plueschke K, McGettigan P, *et al*. Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation. *BMJ Open* 2018;**8**:e023090. doi:10.1136/bmjopen-2018-023090

ABSTRACT

Objective Electronic healthcare databases (EHDs) are useful tools for drug development and safety evaluation but their heterogeneity of structure, validity and access across Europe complicates the conduct of multidatabase studies. In this paper, we provide insight into available EHDs to support regulatory decisions on medicines.

Methods EHDs were identified from publicly available

Strengths and limitations of this study

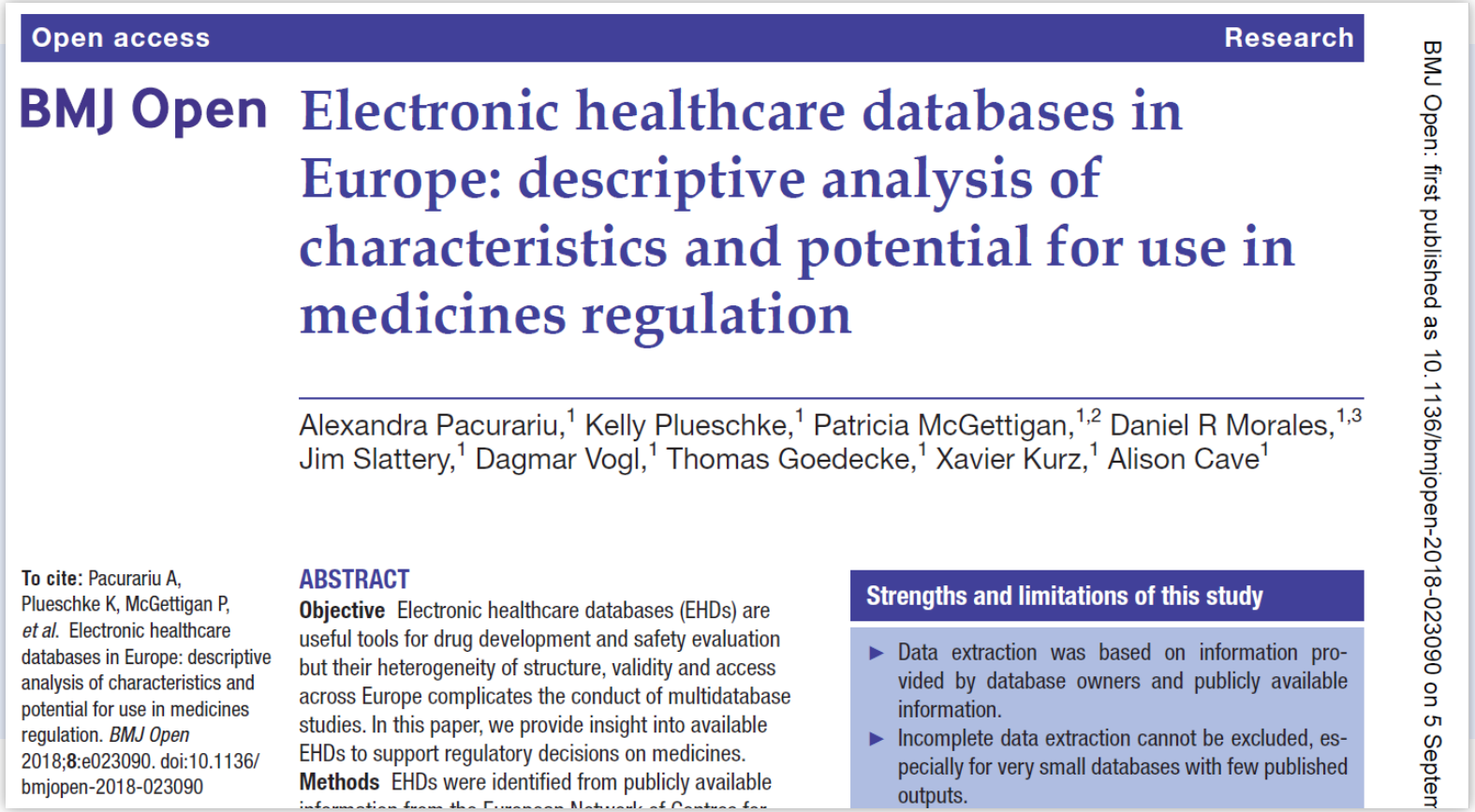
- Data extraction was based on information provided by database owners and publicly available information.
- Incomplete data extraction cannot be excluded, especially for very small databases with few published outputs.

As a first step, we identified existing EHDs in Europe by screening the following sources: the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) resources database,[21] web-based search engines,[22] textbooks on clinical pharmacoepidemiology,[23 24] publicly available inventories created for European Commission-funded research projects and databases used in EMA-funded postauthorisation studies.

As a second step, data sources were included in the inventory based on the following regulatory relevant criteria: the data are available to regulatory authorities or to third parties for research purposes; the database contains information on both drug exposure and health outcomes and is not disease or product specific; there is longitudinal data capture. Provision of relevant data for benefit–risk decision-making was one of the key criteria for selecting studies meeting regulatory requirements.

Prescription-only databases were excluded because they cannot be used for aetiological studies in the absence of the outcome recording. Product-specific or disease-specific registries were considered out of scope as they create cohorts of patients whose entry is defined either by exposure to a product or by occurrence of a disease or health outcome.[25]

Conceptualising European data sources



Open access Research

BMJ Open

Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation

Alexandra Pacurariu,[1] Kelly Plueschke,[1] Patricia McGettigan,[1,2] Daniel R Morales,[1,3] Jim Slattery,[1] Dagmar Vogl,[1] Thomas Goedecke,[1] Xavier Kurz,[1] Alison Cave[1]

To cite: Pacurariu A, Plueschke K, McGettigan P, *et al*. Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation. *BMJ Open* 2018;**8**:e023090. doi:10.1136/bmjopen-2018-023090

ABSTRACT

Objective Electronic healthcare databases (EHDs) are useful tools for drug development and safety evaluation but their heterogeneity of structure, validity and access across Europe complicates the conduct of multidatabase studies. In this paper, we provide insight into available EHDs to support regulatory decisions on medicines.

Methods EHDs were identified from publicly available

Strengths and limitations of this study

- Data extraction was based on information provided by database owners and publicly available information.
- Incomplete data extraction cannot be excluded, especially for very small databases with few published outputs.

As a first step, we identified existing EHDs in Europe by screening the following sources: the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) resources database,[21] web-based search engines,[22] textbooks on clinical pharmacoepidemiology,[23 24] publicly available inventories created for European Commission-funded research projects and databases used in EMA-funded postauthorisation studies.

As a second step, data sources were included in the inventory based on the following regulatory relevant criteria: the data are available to regulatory authorities or to third parties for research purposes; the database contains information on both drug exposure and health outcomes and is not disease or product specific; there is longitudinal data capture. Provision of relevant data for benefit–risk decision-making was one of the key criteria for selecting studies meeting regulatory requirements.

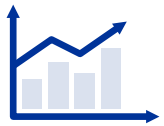
Prescription-only databases were excluded because they cannot be used for aetiological studies in the absence of the outcome recording. Product-specific or disease-specific registries were considered out of scope as they create cohorts of patients whose entry is defined either by exposure to a product or by occurrence of a disease or health outcome.[25]

The data sources were classified in three categories according to their structure, purpose and type of data: electronic medical records, claims databases and healthcare record linkage systems (eg, several databases are linked to form a complete database).

Looking into standard concepts

Electronic record

Any combination of text, graphics, data, audio, pictorial, or other information representation in digital form that is created, modified, maintained, archived, retrieved, or distributed by a computer system. [21 CFR 11.3(b) (6)]



Data model

Unambiguous, formally stated, expression of items, the relationship among items, and the structure of the data in a certain problem area or context of use. A data model uses symbolic conventions agreed to represent content so that content does not lose it's intended meaning

Database

A collection of data or information, typically organised for ease and speed of search and retrieval

Data source

Source of information collected in the course of a clinical trial, specifically used to differentiate between data as collected versus data that are derived or calculated. NOTE: In CDISC, a metadata attribute defined for each data set variable in the Define.xml document of a Study Data Tabulation Model submission that refers to the source of a variable (e.g., case report form, derived, sponsor defined, patient-reported outcome)

CDISC, Glossary v 15.0

CDISC = Clinical Data Interchange Standards Consortium.

Contents

Background and objectives

- In 2019, the Innovative Medicines Initiative (IMI) funded the ConcePTION project with the shared vision that there is a societal obligation to radically and rapidly reduce uncertainty about the safety of medication use in pregnancy and breastfeeding
- A qualitative study was conducted **to identify a conceptual framework** to represent population-based data sources for multidatabase studies—ultimately to make sure they could be faithfully represented into a suitably designed common data model
- For our purposes, the former objective is of importance

Methods

Twenty organisations with access to data were included based on (1) their experience in pregnancy studies and/or in multidatabase studies in pharmacoepidemiology, (2) the data source to which they had access, and (3) their inclusion in the ENCePP network. A systematic characterisation of the data sources was conducted. First, the full data model and data dictionary of data tables accessed by each organisation was requested, in original format. Second, pairs of researchers from the coordinating team reviewed the data dictionaries. Third, each organisation was asked to produce in written form the answers to a fixed set of questions for each table. Fourth, 120-minute interviews were conducted. Fifth, after rounds of revisions, the interview answer document was finalised. Sixth, content analysis was conducted on the interview answer documents: data were extracted in tabular form and iteratively reviewed by coauthors, leading to identification of latent concepts.

Thurin et al,
manuscript in preparation

Methods

1. What **triggers the creation of a record** of the **table**?
2. Is the table collected for all the **population** in your **database**, or only for a **subpopulation**
3. Can you comment on the completeness and quality of the table? If you don't have formal measurements, feel free to convey the assumptions you commonly make
4. What is the time span of the table, how often is it refreshed, and what is the lag time between data entry and the time when data are available for research?
5. Include other comments you may want to share about this table
6. Fill out the table below with the names of the variables in this table (as listed in the data dictionary) that you plan to map to the ConcePTION common data model, with a description in English of the meaning of the variable, the name of the classification used (e.g., CIM10, ATC, or national/local) or the description in English of the data dictionary if a small number of values are included in the dictionary, and any comment you want to share about the variable (e.g., when it is missing or miscoded or when it's content is unreliable)

Thurin et al,
manuscript in preparation

Results: latent concepts

A **data bank** is a collection of structured electronic healthcare data, organised in one or several data tables, about a specific **underlying population,** mandated and sustained by an organisation, the **originator** of the data bank (e.g., the healthcare payer). The **record prompt** of the data bank is the class of events that prompts the generation of a record in a data bank. A **data source** is a collection of data banks having the same underlying population, or overlapping populations, that can be linked to one another at an individual level. A **data access provider (DAP)** for the data source is an institution with authorisation, capacity, and expertise to process its data and interpret results.

Thurin et al,
manuscript in preparation

Results: families of data banks

Thurin et al,
manuscript in preparation

Data bank (number of data sources including it)	Originator of the data bank	Organisations that collect data	Prompts for records in the family of data banks	Typical content	Less common content
Hospital administrative records (16)	Healthcare payer (*)	Healthcare service providers: hospitals	Discharge from a hospitalisation. In some data banks, a specialist encounter would also prompt a record. In some data banks, access to emergency room will also prompt records.	Diagnosis/signs/symptoms observed during hospitalisation; main diagnosis that led to or was developed during the hospitalisation; administration of procedures; specialty of the admission/discharge ward; outcome	Socioeconomic status, results from diagnostic tests, flag of whether the hospitalisation was due to an emergency or was scheduled, flag of whether the hospitalisation included an overnight stay; name of the hospital
Primary care medical records (5)	Network of primary care practices, possibly mandated and/or funded by the healthcare payer or regulatory authority	Primary care practices	Contact between patients and their primary care practice. In some data banks contact must be a face-to-face contact. In others, telephone or online contacts also prompt record creation. Some DAPs (e.g., British DAPs) explained that in their healthcare system the primary care practice does receive automatic information from hospitals when their patients are discharged (e.g., the discharge letter), but this does not in itself prompt a record in the data bank.	Registration with the practice; diagnosis/signs/symptoms; prescription of diagnostic tests; prescription of medicines	Socioeconomic status, results from diagnostic tests; administration of vaccines; dosing regimen of prescribed medicines; indication for a prescribed medicine; results from diagnostic tests; referrals to a specialist or to a hospital
Pharmacy dispensings records (15)	Healthcare payer (*)	Healthcare service providers: pharmacies	Dispensings of a medicinal product admitted to reimbursement by a healthcare payer. According to the data bank, a record may be prompted by community pharmacies, hospital pharmacies, or both. In most data banks, the prompt is limited to dispensings for outpatient use; in other data banks, dispensings for outpatient administration by a healthcare professional (e.g., infusions) also prompt records. In rare cases, dispensings for inpatient use also prompt records.	National code of the medicinal product, ATC code and amount dispensed	Link to the prescription, batch number, condition of pregnancy of the patient, batch number
Birth registry (12)	Public health authority, statistical authority	Hospitals or midwives or children health services	Live births observed, according to the data bank, in hospital or at the first visit of the child. If the prompt is delivery in hospital, stillbirths also prompt a record. If this is the case, the distinction between stillbirth and spontaneous abortion is based on the gestational age at the moment of labour inception and may vary between 20 and 24 weeks, according to the data bank.	Information on the mother, on the pregnancy, on the delivery, on the child(ren)	Information on the father
Induced terminations registry (4)	Public health authority, statistical authority, authority allowing terminations	Hospitals executing the procedure or practices authorising the procedure	Request or execution of an induced termination. The record may be anonymous.	Information on the circumstances of the termination, on the pregnancy, and on the woman	Link to the corresponding hospital discharge record
Congenital anomaly registry (10)	Public health authority and/or research centre	Hospitals of birth or other healthcare professionals involved in the delivery	Recording of a congenital anomaly. The specific prompt depends on the regulation of the EUROCAT centre: it may happen only at birth or during a follow-up of several years. In some data banks, a foetal death with an anomaly would also prompt records.	EUROCAT core variables	Other EUROCAT variables
Inhabitant registry (3)	Civic authority (national or regional)	Civic offices	Immigration/emigration in a country or region.	Immigration/emigration date, birth date	Date of death, address, area of residence
Registration with healthcare system (6)	Healthcare payer (*)	Healthcare service office	Registration with a healthcare system; in some data banks, according to the structure of the healthcare system, this corresponds to association with a primary care physician and/or to registration with a health insurance provider.	Date of registration	Date of de-registration, primary care physician, name of the insurance, address, area of residence
Exemptions from copayment (6)	Healthcare payer (*)	Healthcare service office	Some healthcare payers admit exemptions from copayment due to some health conditions, and a record is prompted when this exemption occurs.	Cause for exemption	
Death registry (6)	Public health authority	Public health authority	Recording of the cause of death.	Principal cause of death	Secondary causes of death, manner of death (e.g., natural, violence, accident, suicide)

* *Single payer* refers to a health system that is financed by a single entity, often a government, e.g. Italy, Denmark, Norway. *Multiple payer* refers to a health system that is financed through more than a single entity, one of which may include government, e.g. Germany, Netherlands

Results: families of data banks

Thurin et al,
manuscript in preparation

Family of data banks	Originator of the data bank	Organisations that collect data	Prompts for records in the family of data banks
Hospital administrative records	Healthcare payer	Healthcare service providers: hospitals	Discharge from a hospitalisation. In some data banks, a specialist encounter would also prompt a record. In some data banks, access to an emergency room will also prompt records.

Results: families of data banks

Thurin et al,
manuscript in preparation

Family of data banks	Originator of the data bank	Organisations that collect data	Prompts for records in the family of data banks
Hospital administrative records	Healthcare payer	Healthcare service providers: hospitals	Discharge from a hospitalisation. In some data banks, a specialist encounter would also prompt a record. In some data banks, access to an emergency room will also prompt records.

Family of data banks	Originator of the data bank	Organisations that collect data	Prompts for records in the family of data banks
Primary care medical records	Network of primary care practices, possibly mandated and/or funded by the healthcare payer or regulatory authority	Primary care practices	Contact between patients and their primary care practice. In some data banks contact must be a face-to-face contact. In others, telephone or online contacts also prompt some records.

Results: families of data banks

Thurin et al,
manuscript in preparation

Family of data banks	Originator of the data bank	Organisations that collect data	Prompts for records in the family of data banks
Hospital administrative records	Healthcare payer	Healthcare service providers: hospitals	Discharge from a hospitalisation. In some data banks, a specialist encounter would also prompt a record. In some data banks, access to an emergency room will also prompt records.

Family of data banks	Originator of the data bank	Organisations that collect data	Prompts for records in the family of data banks
Primary care medical records	Network of primary care practices, possibly mandated and/or funded by the healthcare payer or regulatory authority	Primary care practices	Contact between patients and their primary care practice. In some data banks contact must be a face-to-face contact. In others, telephone or online contacts also prompt some records.

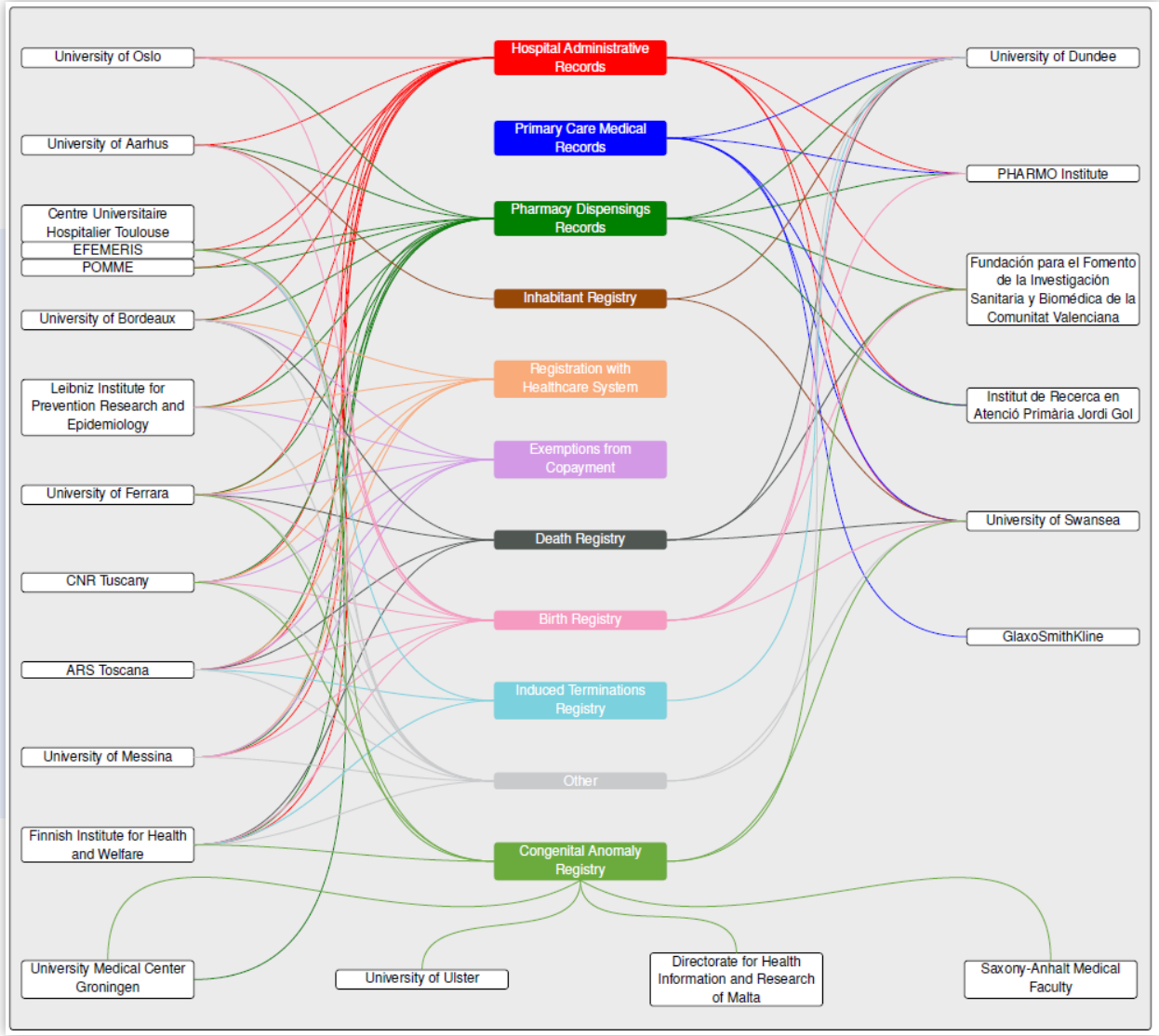
Family of data banks	Originator of the data bank	Organisations that collect data	Prompts for records in the family of data banks
Birth registry	Public health authority, statistical authority	Hospitals or midwives or children health services	Live births observed, according to the data bank, in hospital or at the first visit of the child. If the prompt is delivery in hospital, stillbirths also prompt a record.

Results: data banks included in each data source

4 data sources having 1 data bank

16 data sources having > 1 data banks

Thurin et al,
manuscript in preparation



Contents

Institution

Data sources having 1 data bank

- Uncommon
- The originator of such a data bank may not be among the institutions that provide access to the data source for regulatory purposes

Data sources having > 1 data bank

- These data sources exist because at least one institution has authorisation/capacity/expertise to link the data banks with each other at the level of individual patients
- Often the originators of such data banks are not among the institutions that provide access to the data source for regulatory purposes

Bottom line: access to a data source is often mediated by a DAP that is not the originator of any of its data banks

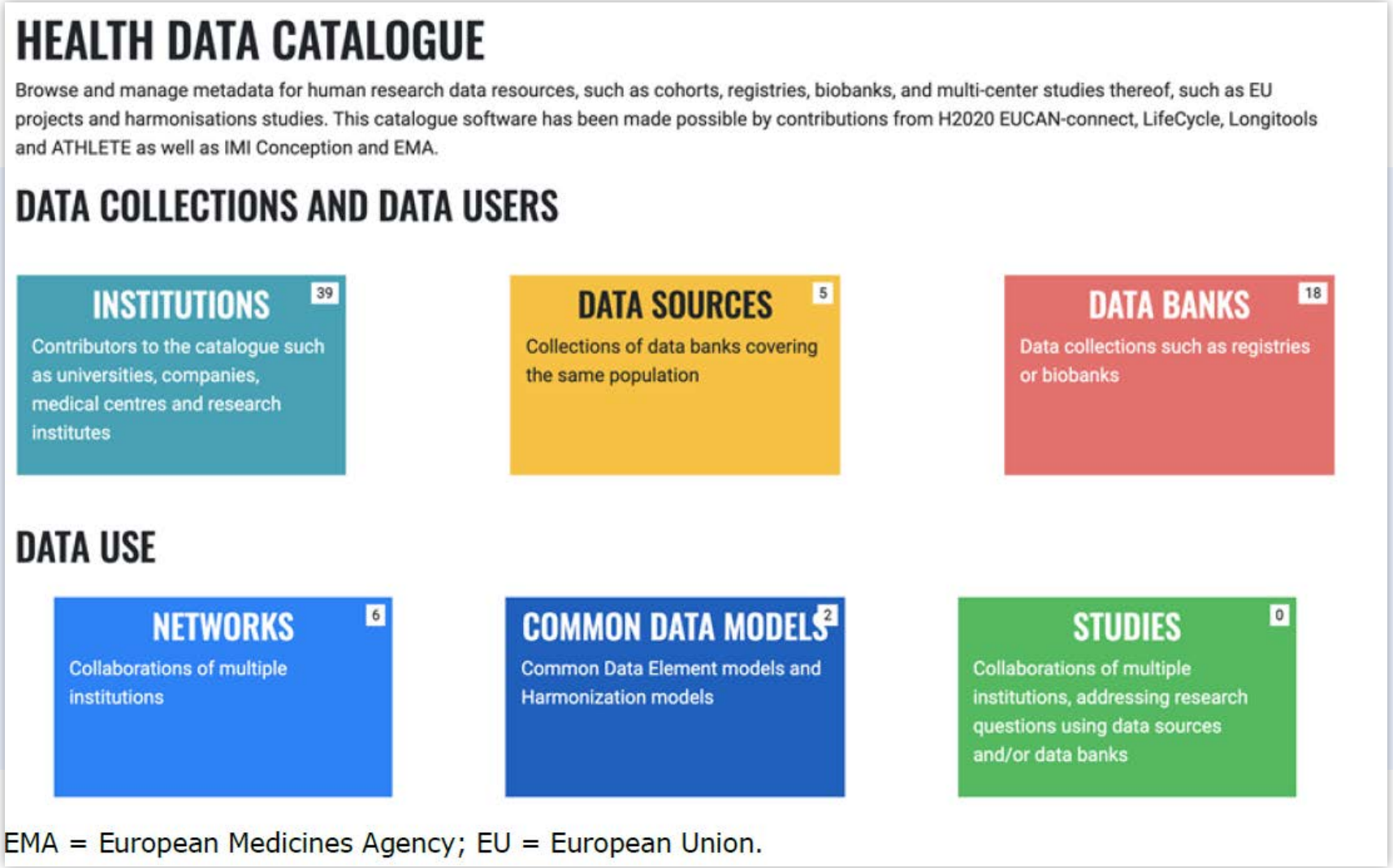
Type of access

Access is regulated

- By law (e.g., GDPR), and
- By agreements/regulations between originator and DAP

As a result, access may be based on

- Institutional purposes
- Licencing agreements
- A protocol



EMA = European Medicines Agency; EU = European Union.

Acknowledgements

Nicolas Thurin
Romin Pajouheshnia
Giuseppe Roberto
Giulia Hyeraci
Claudia Bartolini
Olga Paoletti
Marieke J. Hollestelle
Cailtin N. Dodd
Morris Swertz
Hedvig Marie Egeland Nordeng
Helle Wallach Kildemoes
Vera Ehrenstein
Elena Dudukina
Thomas MacDonald
Giorgia De Paoli
Maria Loane
Christine Damase-Michel
Anna-Belle Beau
Cecile Droz
Regis Lassalle
Jorieke van Kammen-Bergman
Ron Herings
Karin Swart-Polinder
Tania Schink
Clara Cavero
Laia Barrachina Bonet
Ainhoa Gómez Lumbreras
Maria Giner-Soriano
Maria Aragón
Amanda Julie Neville
Annarita Armaroli
Aurora Puccini
Anna Pierini
Valentina Ientile
Janet Sultana
Gianluca Trifirò
Miriam Gatt
Anke Rissmann
Maarit Leinonen
Visa Martikainen
Susan Jordan
Daniel Thayer
Ieuan Scanlon
Betsy Georgiou
Marianne Cunnington
Miriam C.J. Sturkenboom

Thank you!

For any question on this presentation, please contact: Malgorzata.Durka-Grabowska@ema.europa.eu
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