IMI2 821520 - ConcePTION

ConcePTION

WP7 – Information and data governance, ethics, technology, data catalogue and quality support

D7.1 User requirements and metadata model for the FAIR data catalogue from WP1, 2 &7 - task 7.4

Lead contributor	Morris Swertz, Fernanda de Andrade (17 - UMCG)
	m.a.swertz@rug.nl
Other contributors	WP1, Maria Loane, Joanne Given, Florence Coste, Helen Dolk WP2 David Lewis, Eugene van Puijenbroek, Bernke te Winkel, Laura Yates WP7 , Caitlin Dodd, Marianne Cunnington, Rosa Gini, Petr Holub Miriam Sturkenboom,
	17. University Medical Center Groningen 1. University Medical Center Utrecht 41. GlaxoSmithKline 4. BBMRI-ERIC 10. Agenzia Regional di Sanita Tuscany 2. Ulster University 9. Lareb 39. Sanofi 38. Novartis 13. University of KwaZulu-Natal

Document History

Version	Date	Description	Non-contributor reviewers (if applicable)
V0.1	10 Sep 2019	First Draft	Morris Swertz, Fernanda de Andrade
V0.1	19 Nov 2019	Comments	Petr Holub, Miriam Sturkenboom, Rosa Gini, Marianne Cunnington
V0.2	06 Dec 2019	Second Draft	Morris Swertz, Fernanda de Andrade
V0.9	20 Mar 2020	Final Version	Morris Swertz, Fernanda de Andrade, Marianne Cunnington
V1.0	26 Mar 2020	Addition of abbreviations, lay out	Miriam Sturkenboom

Table of Contents

Abbreviations

ADVANCE	Accelerated development of vaccine benefit-risk collaboration in Europe
ADOPT	implementAtion anD OPeration of the gateway for healTh into BBMRI-ERIC
ARS	Agenzia Regionale di Sanita Toscana
BBMRI-ERIC:	European research infrastructure for biobanking
BioMedBridges	BioMedBridges forms a cluster of the biomedical sciences research infrastructures (BMS RIs) and constructs the data and service bridges needed to connect them
BIOSHARE	BioShare is a web-based platform for sharing biospecimens and/or datasets with others in the research community.
CINECA	Common Infrastructure for National Cohorts in Europe, Canada, and Africa
DAP	Data Access Provider
EHDEN	European Health Data Evidence Network
ENCePP	European Network of Centers of Pharmacoepidemiology & Pharmacovigilance
EU	European Union
EUCAN-connect	EUCAN-Connect is a federated FAIR platform enabling large-scale analysis of high-value cohort data connecting Europe and Canada in personalized health. It is a research project funded by the European Commission for five years (2019-2023)
EUROCAT	European network of population-based registries for the epidemiological surveillance of congenital anomalies
ETL	Extract Transformation Load
FAIR:	Findable Accessible Interoperable & Reusable
GDPR	General Data Protection Regulation
GSK	Glaxo Smith Kline
J&J	Johnson & Johnson
Lifecycle	innovative research on the role of novel integrated markers of early-life stressors that influence health across the lifecycle using an open and long-term network of European cohorts that started during pregnancy or childhood
IMI:	Innovative Medicines Initiative
Recap	Research on European Children and Adults born Preterm
UMCG	University Medical Centre Groningen
UMCU	University Medical Centre Utrecht
WP	Work package

Summary

The objective of this report is to summarize user requirements and metadata needs as the basis for the prototype of the FAIR data catalogue for the IMI ConcePTION project, (FAIR data catalogue is Deliverable 7.6).

The objective of the data catalogue is to collect summary level metadata on relevant data sources that enable ConcePTION partners to identify the relevant data sources for the conduct of studies to generate evidence on the safety of medicines in pregnancy (i.e. data source catalogue) and lactation. In addition, the catalogue aims to enable ConcePTION partners to manage and share detailed metadata on the individual data items, their harmonisations and their use in ConcePTION studies (i.e. study and data item catalogue).

To produce these requirements many interviews and brainstorming sessions were organized by and with WPs 1, 2 and 7. In addition desktop research was performed where existing catalogue solutions were compared and documented, and prototype systems developed to give direction of the catalogue development. The results in this report are an overview of the main catalogue uses, expectations on how the catalogue system should behave in terms of search and presentation of data characterisation/ data quality results, decisions on the main modules, and preliminary data models underlying the catalogue system.

The main catalogue modules are planned to be:

- **Collection and data provider/data source discovery service**, providing essential metadata of each data source source with specific extensions for WP1, WP2 and WP7 to understand whether essential elements for their studies are captured.

- **Variable and harmonization description service for studies** (data dictionary, ETL design and pseudo scripts, governance/ethical-legal documents, common data models), to enable detailed description and provenance of each data dictionary and common data model used within the project, as coordinated by WP7.

- And in the future: space and facility to query selected data characterization results generated by WP7.

To maximize 'FAIR'-ness and re-use of existing principles and tools, the catalogue will be designed as an extension of the BBMRI-ERIC Directory (which will maximize Findability). This also facilitates harmonization of contents with existing large EU initiatives in cohort data sharing (which will maximize Interoperability and reuse potential). Therefore, the software basis for the catalogue will be MOLGENIS (a free and open source software with permissive license). The catalogue will also interface to BBMRI-ERIC secure negotiation service to arrange data access for distributed analyses, and a separate private and secure site for simple analysis of data characterisation tables (quality indicators and simple analyses such as prevalence of key variables), which can be used in feasibility studies following user authentication. The catalogue should comply with General Data Protection Regulation (GDPR) where applicable (e.g., substantial parts of the catalogue will contain highly-aggregate non-personal data, but the catalogue will contain some contact information for data sources, and authorisation information for users accessing the catalogue).

Introduction

Motivation

The objective of the data catalogue is to serve the needs of ConcePTION partners to identify relevant data sources for the conduct of studies to generate evidence on the safety of medicines in pregnancy and lactation. It will also allow ConcePTION partners to manage and share metadata on the individual data items and their data harmonisation as used in ConcePTION studies. With longer term sustainability in mind this is being designed with the potential to extend availability of the catalogue to researchers beyond the ConcePTION network.

In particular:

For Task 1.1. the objective of the catalogue is to identify the healthcare data sources that can be used for medication utilisation and medication safety studies in pregnancy by updating and transforming the existing EUROmediSAFE[1] inventory into a searchable form, focusing primarily on existing and new sources in the EU and adding available non-EU data sources.

For Task 2.2 the objective is to determine the structure of a meta data catalogue for industry and publicly held data sources, retrospectively and prospectively collecting data on pregnancy exposures to medication and outcomes (spontaneous reports and pregnancy cohorts). The structure will be reflected through a survey/questionnaire that can subsequently be sent to relevant data access providers to collect their metadata.

For Task 7.4 the objective is to create a FAIR data source catalogue with specific search features that will be defined based on user requirements. The catalogue will use standard metadata models to make data resources findable, understandable, and provide access control for researchers. Metadata is "data that provides information about other data". The metadata will define the resource as a whole, the data sets, data set formats, key variables and their provenance. The catalogue will also interface with a secure negotiation service to arrange data access, and a separate private site for simple analysis of data characterisation tables, which can be used in feasibility services following authentication. The catalogue will initially be filled for data sources relevant to WP1 and 2 research needs. Specifically, this will be the initial 21 DAPs involved in ConcePTION and potentially eligible for WP1 demonstration projects. The total number of data sources for WP2 still has to be confirmed.

Task 7.5 focuses on the creation of remote operating platform and common data models for analysing data in a distributed fashion, while securely sharing results of analyses. Task 7.5 aims to set up the workflow and architecture to make this happen. WP7 is tasked with the data harmonization and characterization of data sources for use in WP1 and WP2. WP7 and Data Access Providers (DAPs) will perform harmonization in two steps: first structural/semantic harmonization will be conducted by the DAPs locally by extracting data required for ConcePTION studies, and transforming the content of that extraction into a low level common data model that aims to do structural harmonization. The data dictionary of original data and the ETL scripts used for harmonization will need to be uploaded to the catalogue for transparency. The Catalogue will also host the results of the semantic harmonization steps data characterisation and enable those results to be queried to identify suitable data sources for planned research.

[1] http://euromedicat.eu/EUROmediSAFE/EUROmediSAFEInventory

Focus on metadata and characterisation of output

The catalogue will only include metadata That means no sensitive, individual level data from study participants will be stored in the catalogue, removing all GDPR considerations except those of the usernames/logins of the ConcePTION partners that will access the catalogue.

Types of metadata include descriptive metadata, structural metadata, administrative metadata (https://marciazeng.slis.kent.edu/metadatabasics/types.htm). In addition, the catalogue will capture characterisation of output, i.e. Characterisation of output is produced through the process of data characterization, which requires descriptive analysis of the data resulting in parameters that effectively describe the characteristics and behaviour of particular data elements. The data characterisation will allow a simple summary of availability of key variables of interest as well as key indicators of data quality such as variable completeness. For further details see Data Characterisation protocol (Appendix 1).

Methods

The following methods were used to collect detailed requirements and to produce initial versions of the catalogue metadata model, which is the content structure for the foundation of the catalogue.

Interviews and group brain-storms

Members of WP1, WP2 and WP7 embarked on separate and combined brainstorms. For WP7 this included discussion between representatives of ARS, UMCU, GSK, J&J, UMCG and BBMRI-ERIC. These discussions revealed the specific requirements of WP7. Task 1.1 produced a “FAIR data source catalogue: Inventory search features and data quality requirements” protocol (Appendix 2) and an “Outline protocol for demonstration projects” which were circulated to all WP1 Partners, and Third Parties for comments, as well as to WP7 to inform the development of the catalogue. WP2 produced a questionnaire format that informed the data to be collected from WP2 data partners to fill the catalogue.

Metadata Catalogue Format

Catalogue user requirements were provided by WP1 and WP2 separately. For WP1 this was provided as a document summarising the query reports that would be run against the catalogue to determine which data sources are fit for purposes to answer specific research questions (Appendix 2). Example queries were provided for medication utilisation and medication safety studies.

For WP2 this took the form of a questionnaire that would be completed by potential DAPs and was stratified according to whether the data were spontaneous report collections or pregnancy research cohorts.

WP7 reviewed these documents along with example catalogues from other large European projects, specifically EnCePP and EHDEN. To align with metadata collection processes across WP1 and WP2 and to align with other catalogue approaches WP7 proposed:

- Using a questionnaire-based approach to collect metadata from DAPs for WP1 and WP2 to allow a direct means of populating and updating the catalogue from DAPs rather than relying on interpretation of publicly available information
- To meet WP1 query requirements, extend the catalogue to capture key elements of data characterisation (e.g. quality indicators and benchmarking summary statistics) that can be linked to the metadata being queried.

- Focus the initial prototype on the 20-30 DAPDAPs who are partners or third parties attached to ConcePTION and extend to other data sources (as specified in Task 1.1) following optimisation of the information collection process and catalogue design.
- Develop a dashboard for visualisation of queries from WP1 and WP2. See Appendix 2 for example of the queries from WP1 that can be considered for the application of visualisation tools.

Data characterisation

To enable data quality indicators to be developed, and to meet the query needs of WP1, a data characterisation approach was developed in **Task 7.6**.

WP1, Task 1.1 produced draft demonstration protocol templates which include a list of required information items as well as a query document which describes example queries they would like to be able to make against the catalogue. Similarly, required items were defined through Task 2.3 from WP2. WP7 translated this into a data characterisation protocol based on previous models of data quality and indicator reporting from networks such as EUROCAT (and EUROlinkCAT and EUROmediCAT), OMOP, EHDEN, ADVANCE and Sentinel, but reflecting the specificity of medication safety studies in pregnancy. Within the catalogue there will be the ability to link the metadata queries to outputs from the data characterisation.

Literature research into existing solutions

We took a pragmatic approach to researching the existing solutions focusing on existing catalogue solutions that might meet ConcePTION needs. In particular, we studied the MIABIS "Minimum Information About Biobanks Information System" and past efforts spent in this research domain on EUROmediSAFE. In addition, we have investigated past cataloguing efforts in general EU projects such as BIOSHARE, ADOPT, BioMedBridges, CINECA as well as exploring the potential to collaborate with currently ongoing projects in the area of pregnancy, birth and child cohort research such as LifeCycle, Recap, EUCAN-connect.

Following development of a prototype catalogue for population-based data based upon the EUROmediSAFE inventory, WP7 reviewed existing metadata catalogues or or those under development or established for observational health care data. Based on expert knowledge within the consortium the review focused specifically on the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) resources database (http://www.encepp.eu/encepp/resourcesDatabase.jsp) as well as the datasource catalogue currently under development in the European Health Data Evidence Network (EHDEN) project (https://docs.google.com/spreadsheets/d/1tS0UaYfd9FOITgoylady4lZjQ5yYOqLvzDhxbvV89GI/edit#gid=0) as these reflect data types most similar to those covered by WP1 within IMI ConcePTION.

The conclusion of this research is that there is a huge potential and willingness to combine common components from these projects into one open source solution maximising re-use potential, which we plan to do moving forward. However, there are also specific needs for ConcePTION that need to be accommodated, specifically around the core data variables and data source heterogeneity. Therefore, a modular approach is advised which can reuse the existing data models and systems as is, and then add ConcePTION specific extensions.

Prototyping

Finally, we developed early prototypes to gain insight if the structure and metadata collection approach was practicable. The catalogue will be implemented using the MOLGENIS open source scientific data platform because that is also behind the BBMRI-ERIC directory. A specific feature of

this platform is that it allows rapid prototyping using data models. Therefore, we have loaded draft data models into preliminary prototype systems using MOLGENIS to enable partners to experiment with the system. These models can be found in the appendices. Inputs into the development of prototypes included:

- EUROmediSAFE inventory
- WP2 extension questionnaire as provided by Eugene Van Puijenbroek of WP2
- WP1 extension questionnaire provided by Caitlin Dodd of WP7 following close collaboration with WP1
- BBMRI-ERIC directory
- LifeCycle variable and harmonization catalogue

For the catalogue prototype, user interfaces were not optimised through major software developments.

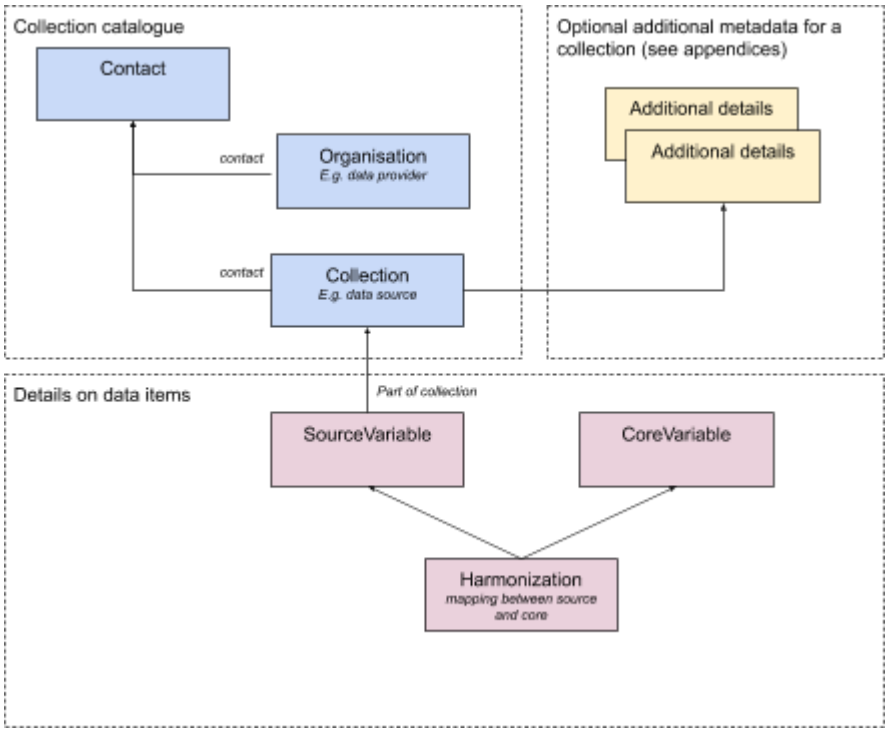
Results

The following modules have been identified, decisions on their contents made, and draft data models for populating these have been created as part of the prototyping effort. There are three main components:

- summary level information per data provider in the 'collection catalogue' (blue, appendix 3)
- optional additional metadata on these data as collected by WP1, WP2 (yellow, appendix 4,5)
- detailed information on data items/data harmonisations efforts e.g. data dictionary collected by WP7 from DAPs (red, appendix 5).

Summary figure and summary descriptions below. Details of the data model can be found in appendix 2-5.

Figure: conceptual overview of catalogue data model



Data source / collection description module

The first level requirements relate to the need to identify relevant data sources or collections that are held by organizations. Task 1.1 provided WP7 with a comprehensive list of European data sources that could potentially be used in ConcePTION. The requirements for this module are summarized in Appendix 2 "collection/data source search requirements".

The type of data sources of interest to ConcePTION include

- pregnancy exposure registries
- disease registries
- population-based record linkage systems
- healthcare database (medical records, administrative claims databases)
- surveillance systems
- educational databases
- birth research cohorts.
- pharmacovigilance databases

Characterization indicators

- Availability of data elements to identify pregnancies and pregnancy timing
- Availability of data elements to identify maternal medication exposure in general and descriptors of specific predefined exposures.
- Availability of outcome data in general (pregnancy outcome, foetal outcome, at birth or later during childhood) and descriptors of specific predefined outcomes.
- Availability of information on confounders (maternal disease, smoking status, BMI, folic acid use etc).
- Indicators of data quality

The characterization indicators are outlined in Appendix 2
We shall only collect high-level aggregate statistical descriptors, to prevent disclosure of potentially privacy sensitive data (i.e. personal identifiable information).

To minimise operational complexity for this module, the BBMRI-ERIC Directory was chosen as a back-bone structure. This enables data exchange with the international catalogue of biobanks (biomaterial collections related to specific disease collections or genome wide association study data), to enable integration into the data request systems of BBMRI-ERIC.

In addition, WP1 and 7 have agreed an approach and structure to transform key elements of the EuroMediSAFE inventory into a searchable catalogue.

Variable and harmonisation description module ('data dictionaries')

Once individual datasets have been identified by the WP1 Demonstration Project leaders, there will be a need to precisely define the contents of those datasets and to assess if the data are suitable for particular research questions. In addition, there is a need to define core data elements/variables that are essential to provide evidence in the studies of WP1 and WP2. In particular, the catalogue should be able to represent the common models as defined in Task 7.5. And finally, there is a need to facilitate data harmonization and integration in the context of pooled analysis of multiple data sources.

Therefore, this module will contain

- precise technical definition of individual variables through their data dictionaries of the datasets involved, including details on types, code systems/vocabularies/ontologies used.
- Similarly, the core variables can be defined either individually or in terms of the individual component elements (i.e. a core variable may be defined as an algorithm combining individual element types). Of specific relevance, we will also load the OMOP common data dictionary into this part of the catalogue to explore to what extent OMOP can provide common model for data harmonization and integration.
- Finally, a system to describe the mappings between the original data source data dictionaries and the core variables used in research analysis will be provided, including ability to document the data harmonisation process across multiple data sources and to record the statistical transformation code used (in e.g. R, SAS, STATA).

To that purpose we request the following possibilities in the catalogue:

Data dictionary space

As part of the process to map the local data to any of the ConcepTION common data models, the data dictionaries of the local data have been requested in task 7.5. WP7 would like a space in the catalogue to upload these data dictionaries for transparency, and as a basis to understand the

extraction transform and load (ETL) processes.

ETL script space

DAPs that participate in ConcePTION will create an ETL script to transform the format of their local data into the ConcePTION CDM. WP7 requests a space on the catalogue to store the ETL script in its original format and the pseudocode (to be available in Year 2).

Common data model space

ConcePTION will have a suite of common data models, geared to the different types of data sources. WP7 would like the catalogue to have a space to store the different CDMs, and the versions (to be available in Year 2).

Reusing Lifecycle data model

In a previous project, EU LifeCycle, the catalogue development has developed a suitable catalogue system for all of above. This work will therefore be the basis for this module of the catalogue.

GDPR Compliance

The proposedcatalogue structure and contents indicate that it will be possible to build the catalogue using only highly aggregate data and thus de facto non-individual level-personal data. This implies that the catalogue contents is not subject to current data protection regulations. However, the catalogue will collect and link to publically available contact information associated with the DAPs and their data sources, as well as only essential information about users to enable authorized access. A data privacy statement will therefore be agreed by the DAPs and partners who are participating and contributing to the catalogue.

Access Negotiation

After the candidate data sources are identified by the Demonstration Project Leaders in WP1 and WP2, WP7 will negotiate access with the relevant DAP. Hence the catalogue will need to interface a secure access negotiation service, which will also ensure authentication and appropriate authorization of the parties involved in the negotiation for the ConcePTION project.
Practical implementation of the access negotiation will use BBMRI-ERIC Negotiator and BBMRI-ERIC Authentication and Authorization Infrastructure (AAI). The AAI service is already in use for the task management system used to control the process of data collection and updates in the ConcePTION project, and hence the data source representatives are already registered in it.

Conclusion

This report provides the building blocks to advance deliverable D7.6 Prototype of FAIR catalogue. In particular, all of above has resulted in a granular specification of the data model used for the subsequent catalogue deliverable, i.e., prototype. Find details of the core catalogue model as well as extensions serving WP1 and WP2 specific needs in the Appendices below.

Appendix 1:

Data Characterisation Protocol

ConcePTION Data Characterization Indic

Document can be provided upon request.

Appendix 2:

FAIR data source catalogue: Inventory search features and data quality requirements

Produced by WP1 Task 1.1.

The user will have a research question they want to answer relating to medication utilisation or medication safety during pregnancy and lactation. They need to know which databases, or linked databases, containing *individual* data on both mothers and children will allow them to conduct such an investigation. The ideal linked dataset for a medication utilisation/safety study will, for a large unbiased population:

- Identify pregnancies and pregnancy timing
- Provide information on maternal medication exposure and indication
- Provide outcome data (either pregnancy outcome or offspring outcome, at birth or later during childhood)
- Provide information on confounders (maternal disease, smoking status, BMI, folic acid use etc).

The types of data sources which may be useful include pregnancy exposure registries, population-based record linkage surveillance systems, healthcare database (medical records, administrative claims databases, adverse drug reaction databases), purpose-built surveillance systems, educational databases and birth cohorts (this is a preliminary list). For a data source to be included in the catalogue it needs to be possible to use it for a medication utilisation, medication safety study or both – directly or through linkage with other data sources.

In order to answer their research question the user will need certain information items to be available in the database or complex of databases. An information item may be a variable itself, but it may also be possible to derive an information item from a combination of variables. For example, to be able to tell if a medication was used during pregnancy the user would need to know when the pregnancy started and ended. The pregnancy end date would be the date of birth (a variable in a database), or date of other pregnancy outcome, but the pregnancy start date may need to be calculated by subtracting the gestational age from the date of birth.

The information items which would be useful in medication utilisation and medication safety studies are shown in Table 1 of the 'Outline protocol for demonstration projects' document and fall into categories i.e. pregnancy timing, medication exposure, maternal disease/medication indication, outcomes and confounders/covariates.

It should be possible to search on any combination of these in an either/or fashion e.g. (community prescription OR hospital outpatient prescription) AND dose AND …) to find the databases which contain them and to search for any or all of these within the query. It should also be possible to search based on the type of data source for e.g. for a disease registry or birth registry. Example queries are provided below.

The catalogue would provide a report detailing which countries have a data source/linked data source complex which could be used to answer the research question – specifying the data sources. The linkage between databases may be to link mother to child, and/or to link different information sources for the mother, and/or for the child. The report should provide a link to a related file providing more detail of each data source (as above) including the relevant variables (how they are coded + metadata regarding data quality), more detail on the population covered by the database (years, how the population is selected e.g. age, all population or births, accessing specific

type of healthcare, number per year and cumulative number) and the potential for linkage to other data sources.

Example queries to return a list of countries with a data source/linked data source complex which could be used to answer the research question

Medication utilisation query example 1

Research question:
What is the prevalence (% of women) using tamoxifen for breast cancer in the first, second and third trimesters of pregnancy? To answer this question the user will need to know the start and end date of the pregnancy, date(s) and duration of tamoxifen use. Therefore, the catalogue must answer the question: "what data sources can potentially help me in answering this question".

The information items which could be used to answer this question are:

- Pregnancy timing AND Name/code of medication AND Date of prescription/dispensation AND Quantity dispensed AND (Strength OR Dosage instructions).
- Data sources potentially useful would be pregnancy exposure registries, population-based record linkage surveillance systems, Healthcare database (medical records, Administrative claims databases) and purpose-built surveillance systems.
- prescription data, maternity data, medical birth registry, primary care data, Cancer registry.

Medication utilisation query example 2

Research question:
What is the prevalence (% of women) using tricyclic antidepressants for neuropathic pain during pregnancy? To answer this question the user would need to know the start and end date of the pregnancy, date(s) of use of tricyclic antidepressants and the indication for their use.

The information items which could be used to answer this question are: Pregnancy timing AND Name/code of medication AND Date of prescription/dispensation AND (strength of medication OR diagnosis in healthcare database OR co-medications OR prescriber speciality).

Data sources potentially useful would be population-based record linkage surveillance systems and healthcare databases (medical records, Administrative claims databases).

Medication safety query

The users research question: to investigate the safety of antidepressants (e.g. SSRI) medications taken during pregnancy, in relation to risk of adverse perinatal outcomes and later child neurodevelopmental problems. To do this it would be necessary to compare the neurodevelopmental outcomes observed in children exposed during pregnancy (prenatally) to those of children from either the general population and those born to women with the same medical condition (depression) but unexposed to the medication of interest or similar medicines. To answer this question the user would need to know the start and end date of the pregnancy, date(s) of use of antidepressants, perinatal outcomes, child neurodevelopmental outcomes and confounders/covariates which may impact on these outcomes.

Information items needed to answer the research question:

Pregnancy timing AND Name/code of medication AND Date of prescription/dispensation AND (CA diagnosis OR Termination of Pregnancy for Foetal Anomaly) AND (Child referrals to specialists OR

Academic Examination Results OR Childhood prescriptions OR Psychometric measurements OR Health care diagnosis) AND (Maternal age at delivery AND folic acid use AND Breastfeeding AND maternal socioeconomic status AND highest maternal education AND other risk factors).

Data sources potentially useful would be population-based record linkage surveillance systems, healthcare database (medical records, Administrative claims databases, adverse drug reaction databases), purpose-built surveillance systems and educational databases.

Envisioned output

When queried the catalogue should produce a report (formatted as a document) detailing:

1) all countries with a data source/linked data source complex which could be used to answer the research question – specifying the data sources.
2) For each data source:

- Type of data source- Pregnancy exposure registries, Population-based record linkage surveillance systems, Healthcare database (medical records, Administrative claims databases, adverse drug reaction databases), Purpose built surveillance systems, Educational databases and birth cohorts.
- Information items provided

1. Population o Years available (with medication data/linkable to medication data) o Level of data – individual or aggregated o Number per year and cumulative number of women of childbearing age, pregnancies, births or children o If available the number per year and cumulative number of pregnancy, perinatal and childhood outcomes and confounders available for e.g. live births, stillbirths, terminations of pregnancy (elective or for foetal anomaly), spontaneous abortions, major congenital malformations and women breastfeeding. o Potential overlap with other data sources
2. Linkage o Potential for linkage ▪ Previously linked for medication utilisation/safety in pregnancy research. Potentially linkable (Not linkable – NOTE: such data sources should only be included in the catalogue if they can be used on their own (without needing linked to other sources) for medication utilisation/safety in pregnancy research) o Other databases it has been linked to o What variables have been used to link subjects i.e. is it a unique identifier or personal contact details (name/ address etc) o Description of how the mother child link is done o Facility/linkage environment available o Linkage report - % linked/unlinked and reason for non-linkage
3. Data information o Data dictionary/meta data/Variables which correspond to each information item and their coding structure (variable format & variable codes/ values) o How variables were originally collected/defined/details of data entry professional o Source of maternal drug exposure – self-report, prescription, dispensed medication o Data quality/completeness assessment (data sources own) o Language of data source
4. Data access o Cost structure o Access permissions needed/process to gain access o Mean time from application submission to data becoming available for analysis o User guides o collaboration and ownership information
5. Publications written using the data source
6. Contact details

Note: Not all the above will be available for every data source. Data characterisation will also be available for some data sources providing additional information.

Contextual information to include in the catalogue
Information on each country's health system:

- % public/private health care
- Standard of care for pregnancy management (e.g. visits to HCP, ultra-sound procedures)

- Access to elective pregnancy termination
- Delivery statistics (prevalence number/rate, delivery type (c-section))
- Reimbursement of prescription medication (None, All, partially (with patient co-pay), during pregnancy, for certain diseases) as well as contraceptive methods.

Potential sources of information include.

- http://www.euro.who.int/__data/assets/pdf_file/0011/376625/pharmaceutical-reimbursement-eng.pdf
- https://ec.europa.eu/docsroom/documents/7605?locale=en
- https://ppri.goeg.at/
- https://ec.europa.eu/docsroom/documents/7604?locale=en

Appendix 3: Catalogue specification for health data sources aligned between WPs 1 & 7

These sections enable the catalogue to capture the basic information about data sources. This is a loose extension based on MIABIS standard (in particular 'collection' and 'biobank' to ease future sharing of ConcePTION data with BBMRI-ERIC Directory.

Tables include:

- Collection (containing many details on each data source)
- Organisation (containing organisational details of the data source provider)
- Contact (public contact information for accessing the data)

Entity	Attribute	RefEntity	Description
catalogue_organisation	ID		Identifier
catalogue_organisation	Name of organisation		Official organization name (preferable as registered at EU)
catalogue_organisation	Acronym of organisation		Acronym of organization used in EU projects.
catalogue_organisation	Type of organisation	dap_organisation_type	Official classification for EU projects
catalogue_organisation	Description of Other type of organisation		
catalogue_organisation	Location of Organisation	dap_europe	European country where organisation is located.
catalogue_organisation	Dedicated website for the Organisation		
catalogue_organisation	Administrative contact person(s)	catalogue_contact	Administrative contact person (allowed to sign contracts)
catalogue_organisation	Scientific contact person(s)	catalogue_contact	Scientific contact person for pregnancy related studies
catalogue_organisation	Privacy officer(s)	catalogue_contact	Privacy officer for access to data source(s)
catalogue_organisation	Creation date		
catalogue_organisation	Creator of entry (WP1 questionnaire, WP2 Questionnaire, import, manual)	catalogue_creator	
catalogue_contact	Last date Modification		

catalogue_contact	Name of ContactPerson		Contactperson(s) can be Administrative, Scientific or Privacy.
catalogue_contact	Type of contactperson (person can have multiple roles)	catalogue_contacttype	Type of contactperson
catalogue_collection	Email address for contactperson		
catalogue_collection	ID		
catalogue_collection	Name of the datasource		Name of the specific datasource
catalogue_collection	Acronym of the Datasource		Acronym of the Datasource
catalogue_collection	Organisation	catalogue_organisation	
catalogue_collection	Dedicated website for Datasource		
catalogue_collection	Summary description Datasource		Publicly available description of Datasource
catalogue_collection	Datatype	dap_datatype	Type of Datasource
catalogue_collection	Description of 'other' Datatype		Description of 'other' type of Datasource
catalogue_collection	ContactPerson Datasource	catalogue_contact	
catalogue_collection	Data capture process(es)	dap_capture	Data capture process(es) are used in datasource
catalogue_collection	Description of 'other' Data capture process(es)		
catalogue_collection	Update frequency	dap_update	Frequency of source data updates
catalogue_collection	Last update		What is the date of the last update of the datasource
catalogue_collection	start Year		Starting year of coverage datasource
catalogue_collection	end Year		End year of coverage datasource

catalogue_collection	Total number of unique individuals captured	dap_numberpart	Number of unique individuals captured at any time in the data source
catalogue_collection	Number of unique individuals captured lastyear	dap_numberpart	Number of unique individuals captured in the past calendar year in the data source
catalogue_collection	General Variables used in Datasource		
catalogue_collection	Data originates from Country(s)	dap_country	Data originates from which part of the world
catalogue_collection	Data originates from European country(s)	dap_europe	Data originates from which European Economic Area countries
catalogue_collection	Geographic coverage	dap_coverage	Geographic coverage of data
catalogue_collection	Description of 'other' Geographic coverage		
catalogue_collection	Primary language		
catalogue_collection	Admission Criteria	dap_admission	Possible causes for a person to enter/exit Datasource
catalogue_collection	Description of 'other' Admission Criteria		
catalogue_collection	Agegroups	dap_agegroup	Which agegroups are included in datasource
catalogue_collection	Type of care setting(s)	dap_care	Care setting(s) covered by datasource
catalogue_collection	Description of 'other' type of care setting covered		
catalogue_collection	Source of medication information	dap_expoinfo	Source of medication information
catalogue_collection	Medications reimbursment	dap_medication	
catalogue_collection	Contraceptives Reimbursment	dap_contraceptive	
catalogue_collection	Pregnancy Terminations	dap_termination	Subjects in the population covered by this data source access to elective pregnancy

			terminations
catalogue_collection	Source of diagnosis information	dap_outcomeinfo	Source of diagnosis information
catalogue_collection	Coding system used for procedures	dap_codeproces	Coding system used for procedures
catalogue_collection	Description of 'other' coding system used for procedures		
catalogue_collection	Coding system used for drugs	dap_codedrugs	Coding system used for drugs
catalogue_collection	Description of 'other' Coding system used for drugs		
catalogue_collection	Coding system is used for diagnoses	dap_codediagnosis	Coding system is used for diagnoses
catalogue_collection	Description of 'other' Coding system is used for diagnoses		
catalogue_collection	ConcePTION Variables used in Datasource		
catalogue_collection	Type of drug exposure data	dap_drugexpo	Type(s) of data available on drug exposure
catalogue_collection	Description of 'other' type of drug exposure data		
catalogue_collection	Type of drug source data	dap_drugsource	Type(s) of sources on drug data
catalogue_collection	Description of 'other' type of drug source data		
catalogue_collection	Indication for use	dap_yesno	Indication for use recorded (Y/N)
catalogue_collection	Breastfeeding duration	dap_yesno	Breastfeeding duration captured (Y/N)
catalogue_collection	Exclusive breastfeeding	dap_yesno	Exclusive breastfeeding captured (Y/N)
catalogue_collection	Academic Level Child	dap_yesno	Academic performance in children captured (Y/N)

catalogue_collection	Pregnancy Elements	dap_pregelements	Element(s) specific to pregnancy captured in datasource
catalogue_collection	Medical Events	dap_pregevents	Data available on medical events
catalogue_collection	Source capturing Events	dap_eventsources	Source(s) used for capturing medical event(s)
catalogue_collection	Description of 'other' source capturing event		
catalogue_collection	Documentation for datasource		
catalogue_collection	Data dictionary present	dap_yesno	Data dictionary present for Datasource (Y/N)
catalogue_collection	Summary Data dictionary		Summary information on data dictionary of Datasource
catalogue_collection	Location Application Forms		Location for applications forms for access to datasource
catalogue_collection	Publications		Key publications using the datasource
catalogue_collection	Governance		
catalogue_collection	Policy Data Access	dap_yesno	Written policy governing data access to datasource (Y/N)
catalogue_collection	Committee Data Access	dap_yesno	Committee to evaluate requests for data access to datasource (Y/N)
catalogue_collection	Additional information on Subject	dap_yesno	Possibility (both legally and practically) to obtain additional information on a subject (Y/N)
catalogue_collection	Source Additional information	dap_addinfoOptions	Sources to obtain additional information on subjects
catalogue_collection	Description of 'other' Source(s) Additional information		
catalogue_collection	Restriction aggregated data	dap_yesno	Restrictions on sharing of aggregate data with cell counts below a certain number(Y/N)
catalogue_collection	Smallest cell count		

catalogue_collection	Fees on data access	dap_charge	Fee apply for data access to datasource (Y/N/academic fee)
catalogue_collection	Process Time requested data	dap_processTime	Typical time from application submission to data becoming available for analysis
catalogue_collection	Handling Processes		
catalogue_collection	Linkage to other sources	dap_yesno	
catalogue_collection	Linkage source(s)		Previously linked to other data sources (Y/N)
catalogue_collection	Linkage Variables		Variable(s) used to link the data source to other data sources
catalogue_collection	Possible Linkage	dap_link	Possibility to link datasource with other datasources
catalogue_collection	Possible Linkage Sources		Datasources were linkage is possible
catalogue_collection	Mother-child linkage	dap_motherchild	
catalogue_collection	Mother-child linkage other		
catalogue_collection	How is data quality monitored?	dap_quality	Quality monitored from datasource
catalogue_contacttype	Id		
catalogue_creator	id		

Appendix 4: Specification of contents for catalogue from WP2, dealing with pharmacovigilance and exposure cohorts

This section provides catalogue parameters to capture details on pharmacovigilance and exposure. The data model consists of main table 'questionnaire' as well as many coded lookup lists such as 'quality', 'codeDrugs', etc.

entity	label-en	refEntity
dap_questionnaire	id	
dap_questionnaire	Organizational information	
dap_questionnaire	What is the offical name of your organization as used in EU projects?	
dap_questionnaire	What is the acronym of your organization that you use in EU projects?	
dap_questionnaire	Type of organisation	dap_organisation_type
dap_questionnaire	Please specify, if Other is selected	
dap_questionnaire	What country is your organisation located?	dap_europe
dap_questionnaire	In case of a dedicated website for the Organization, please provide the address (URL)	
dap_questionnaire	Who is the administrative contact person (allowed to sign contracts)?	
dap_questionnaire	Please provide email address of the administrative contact person	
dap_questionnaire	Who is the scientific contact person for pregnancy related studies	
dap_questionnaire	Please provide email address of the scientific contact person	
dap_questionnaire	Who is the data privacy officer with whom you liaise for access to the data source?	
dap_questionnaire	Please provide email address of the data privacy officer	
dap_questionnaire	Characteristics Datasource	

dap_questionnaire	What is the name of the specific datasource your organization can access (per datasource, one questionnaire)?	
dap_questionnaire	What is the acronym of the Datasource (if any)?	
dap_questionnaire	In case of a dedicated website for this Datasource please provide the address (URL)	
dap_questionnaire	Is there a publically available Datasource description?	dap_yesno
dap_questionnaire	If Yes, please give summary description or provide the address (URL) to description.	
dap_questionnaire	Is there a contactperson(s) dedicated to this datasource (Other then contactinformation provided for Organization)?	dap_yesno
dap_questionnaire	Name of contact for datasource	
dap_questionnaire	Please provide email address of contact person	
dap_questionnaire	From which country/countries is the data originating?	dap_country
dap_questionnaire	If European Economic Area, please select all applicable EEA countries	dap_europe
dap_questionnaire	What is the primary language used in the data source?	
dap_questionnaire	What is the geographic coverage?	dap_coverage
dap_questionnaire	If other, please specify	
dap_questionnaire	From which age groups do the data come?	dap_agegroup
dap_questionnaire	What is the Datasource type?	dap_datatype
dap_questionnaire	If other, please specify	
dap_questionnaire	What care setting(s) are covered?	dap_care
dap_questionnaire	If other, please specify	
dap_questionnaire	What data capture process(es) are used?	dap_capture
dap_questionnaire	If other, please specify	

dap_questionnaire	What are the possible causes for a person to enter/exit your Datasource?	dap_admission
dap_questionnaire	If other, please specify	
dap_questionnaire	What is the frequency of source data updates?	dap_update
dap_questionnaire	If other, please specify	
dap_questionnaire	What is the date of the last update of the datasource?	
dap_questionnaire	What is the starting year of coverage/data collection?	
dap_questionnaire	What is the end year of coverage/data collection?	
dap_questionnaire	What is the number of unique individuals captured at any time in the data source?	dap_numberpart
dap_questionnaire	What is the number of unique individuals captured in the past calendar year in the data source?	dap_numberpart
dap_questionnaire	What is the source of medication information?	dap_expoinfo
dap_questionnaire	Are medications reimbursed in the population covered by this data source?	dap_medication
dap_questionnaire	Are contraceptives reimbursed in the population covered by this data source?	dap_contraceptive
dap_questionnaire	Do subjects in the population covered by this data source have access to elective pregnancy terminations?	dap_termination
dap_questionnaire	What is the source of diagnosis information?	dap_outcomeinfo
dap_questionnaire	What coding system is used for procedures?	dap_codeproces
dap_questionnaire	If other, please specify	
dap_questionnaire	What coding system is used for drugs?	dap_codedrugs
dap_questionnaire	If other, please specify	
dap_questionnaire	What coding system is used for diagnoses?	dap_codediagnosis

dap_questionnaire	If other, please specify	
dap_questionnaire	Variables used in Datasource	
dap_questionnaire	What data is available on medication exposure?	dap_drugexpo
dap_questionnaire	If other, please specify	
dap_questionnaire	What sources of medication data are captured?	dap_drugsource
dap_questionnaire	If other, please specify	
dap_questionnaire	Is the indication for medication use recorded?	dap_yesno
dap_questionnaire	Is data on breastfeeding duration captured?	dap_yesno
dap_questionnaire	Is data on breastfeeding exclusivity captured?	dap_yesno
dap_questionnaire	Is data on academic performance in children captured?	dap_yesno
dap_questionnaire	What elements specific to pregnancy are captured in the data source?	dap_pregelements
dap_questionnaire	What data is available on medical events?	dap_pregevents
dap_questionnaire	What sources of medical event data are captured?	dap_eventsources
dap_questionnaire	If other, please specify	
dap_questionnaire	Documentation	
dap_questionnaire	Is there a data dictionary available for the datasource?	dap_yesno
dap_questionnaire	If Yes, please give summary info on data dictionary.	
dap_questionnaire	If available upload data dictionary	sys_FileMeta
dap_questionnaire	Please provide some key publications using the data (max 10)	
dap_questionnaire	Governance	
dap_questionnaire	Is there a written policy governing data access?	dap_yesno
dap_questionnaire	If Yes, please upload policy governing data access	sys_FileMeta

dap_questionnaire	Is there a committee to evaluate requests for data access?	dap_yesno
dap_questionnaire	Is it possible (both legally and practically) to obtain additional information on a subject?	dap_yesno
dap_questionnaire	If yes, from what sources can additional information be obtained?	dap_addinfoOptions
dap_questionnaire	If other, please specify	
dap_questionnaire	Is there a restriction on sharing of aggregate data with cell counts below a certain number?	dap_yesno
dap_questionnaire	If yes, what is the smallest cell count which can be shared?	
dap_questionnaire	Is a charge made for data access?	dap_charge
dap_questionnaire	What is the typical time from application submission to data becoming available for analysis?	dap_processTime
dap_questionnaire	Handling Processes	
dap_questionnaire	Has this data source previously been linked on an individual patient basis to other data sources?	dap_yesno
dap_questionnaire	If yes, to which other data sources has the data source been linked?	
dap_questionnaire	What variable(s) was used to link the data source to other data sources?	
dap_questionnaire	Is it possible to link this data source with other data sources to which it has not been linked previously?	dap_link
dap_questionnaire	If yes, to which other data sources is linkage possible?	
dap_questionnaire	How is mother-child linkage conducted	dap_motherchild
dap_questionnaire	If other, please specify	
dap_questionnaire	How is data quality monitored?	dap_quality
dap_quality	id	
dap_link	id	
dap_charge	id	
dap_addinfoOptions	id	
dap_drugexpo	id	

dap_drugsource	id	
dap_yesno	id	
dap_pregelements	id	
dap_pregevents	id	
dap_eventsources	id	
dap_capture	id	
dap_admission	id	
dap_update	id	
dap_numberpart	id	
dap_expoinfo	id	
dap_outcomeinfo	id	
dap_codeproces	id	
dap_codedrugs	id	
dap_codediagnosis	id	
dap_datatype	id	
dap_care	id	
dap_agegroup	id	
dap_coverage	id	
dap_organisation_type	id	
dap_europe	id	
dap_europe	name	
dap_country	id	
dap_medication	id	
dap_contraceptive	id	
dap_termination	id	
dap_processTime	id	
dap_motherchild	id	

Appendix 5: Specification of questionnaire/survey to collect data into the catalogue for WP2 (questionnaire)

This add-on module defines general survey metadata to be collected by WP2. Analogous to previous appendix this consist of one large table 'questionaire' with many small tables to describe coded lookup lists such as 'breast_feed_outcomes', 'long_term_oucomes', etc.

entity	label-en	refEntity
conception_questionnaire	id	
conception_questionnaire	Organizational information	
conception_questionnaire	What is the name of your organisation?	
conception_questionnaire	What is the nature of the organisation?	conception_organisation_type
conception_questionnaire	Please specify, if Other is selected	
conception_questionnaire	What is the acronym	
conception_questionnaire	What type of data do you collect in your organisation?	conception_collection_type
conception_questionnaire	If none of the above, please explain the main objective	
conception_questionnaire	Thank you for paticipation. Do you have feedback?	
conception_questionnaire	Characteristics Organisation	
conception_questionnaire	Characteristics Spontaneous Data Organisation	
conception_questionnaire	Who is the contact person for questions related to the content of the data?	
conception_questionnaire	Please provide the email address of this contact person	
conception_questionnaire	What is the affiliation of this contact person?	
conception_questionnaire	Who is the contact person for technical aspect regarding the data collection?	
conception_questionnaire	Please provide the email address of this contact person	
conception_questionnaire	What is the affiliation of this contact person?	
conception_questionnaire	What is the population covered?	conception_population
conception_questionnaire	What is the starting year of coverage?	

conception_questionnaire	What is the end year of coverage?	
conception_questionnaire	What is the number of reports of the collection?	
conception_questionnaire	What is the number of reports per year?	
conception_questionnaire	Who is reporting to the database?	conception_reporter_type
conception_questionnaire	Please specify, Other reporter	
conception_questionnaire	Is there a dedicated form to collect information about exposures during pregnancy?	conception_yesno
conception_questionnaire	Is there a dedicated form to document exposures during lactation?	conception_yesno
conception_questionnaire	How are pregnancy and breastfeeding related data being used?	conception_data_use
conception_questionnaire	Please specify, if Other is selected	
conception_questionnaire	Characteristics Pregnancy Cohort Data Organisation	
conception_questionnaire	Who is the contact person for questions related to the content of the data?	
conception_questionnaire	Please provide the email address of the contact person	
conception_questionnaire	What is the affiliation of this contact person?	
conception_questionnaire	Who is the contact person for technical aspects regarding the data collection?	
conception_questionnaire	Please provide the email address of this contact person	
conception_questionnaire	What is the affiliation of the contact person?	
conception_questionnaire	Is one of the primary objectives of your data collection the evaluation of the use of medication in pregnancy?	conception_yesno
conception_questionnaire	What is the name of the pregnancy cohort data collection?	
conception_questionnaire	In case of a dedicated website for this cohort, please provide the address (URL)	
conception_questionnaire	What is the nature of the pregnancy cohort?	conception_nature_data
conception_questionnaire	In case you collect information on a specific drug(group) or indication, please specify	
conception_questionnaire	What is the region covered?	conception_population

conception_questionnaire	What is the starting year of coverage?	
conception_questionnaire	What is the end year of coverage?	
conception_questionnaire	What is the total number of inclusions?	
conception_questionnaire	What is the annual number of inclusions?	
conception_questionnaire	Who provides the data?	conception_data_provider
conception_questionnaire	Please specify, Other Data provider	
conception_questionnaire	Is reporting data from the pregnancy cohort to national pharmacovigilance centre mandatory?	conception_yesno
conception_questionnaire	Please specify	
conception_questionnaire	How are pregnancy and breastfeeding related data being used?	conception_data_use
conception_questionnaire	Please specify, if Other is selected	
conception_questionnaire	Variables	
conception_questionnaire	Variables of Spontaneous Data that are collected	
conception_questionnaire	What medicines do you monitor?	conception_exposure_spont
conception_questionnaire	Please specify the specific drugs	
conception_questionnaire	Do you allocate a 'suspect drug'?	conception_suspect_spont
conception_questionnaire	If yes, do you register co-medications?	conception_yesno
conception_questionnaire	Do you use a coding system for classification of drugs?	conception_coding_spont
conception_questionnaire	Please specify which coding system you use (e.g. ATC or other)	
conception_questionnaire	Do you collect information on dosing?	conception_yesno
conception_questionnaire	Do you collect information on timing of exposure?	conception_yesno
conception_questionnaire	Do you only register adverse outcomes?	conception_adv_outcome
conception_questionnaire	What type adverse outcomes do you collect on death?	conception_death_spont
conception_questionnaire	What type of adverse pregnancy outcomes do you collect on pregnancy complications?	conception_preg_compl_spont
conception_questionnaire	What type of adverse pregnancy outcomes do you collect on delivery complications?	conception_delivery_compl_spont
conception_questionnaire	What type of adverse breastfeeding outcomes do you collect?	conception_breastfeed_outc

conception_questionnaire	If other, please specify	
conception_questionnaire	What other types of information do you specifically register for spontaneous reports?	conception_other_information
conception_questionnaire	Do you specifically register information on long term outcomes for spontaneous reports?	conception_longterm_outcomes
conception_questionnaire	Variables Pregnancy Cohort Data	
conception_questionnaire	What is the source of outcome information?	conception_outcome_source
conception_questionnaire	Are there restrictions for inclusion regarding gestational age?	conception_restr_incl
conception_questionnaire	Other , please specify	
conception_questionnaire	How many data collections moments are intended during pregnancy?	
conception_questionnaire	Please specify the time points	
conception_questionnaire	How many data collections moments are intended after pregnancy?	
conception_questionnaire	Please specify the time points	
conception_questionnaire	At what time does the regular follow-up end?	conception_follow_up
conception_questionnaire	Other , please specify	
conception_questionnaire	Do you ask for additional information?	conception_add_info_inquiry
conception_questionnaire	Please specify	
conception_questionnaire	Who provides the additional information?	
conception_questionnaire	Do you have a comparison group?	conception_comparison_group
conception_questionnaire	Please specify	
conception_questionnaire	What is the source of exposure information?	conception_exposure_source
conception_questionnaire	What information do you collect on drug exposure?	conception_info_exposure_drug
conception_questionnaire	What information do you collect on the type of drug?	conception_info_exposure_type
conception_questionnaire	What types of outcome do you collect with regards to pregnancy outcome?	conception_out_type_pregoutc
conception_questionnaire	What types of outcome do you collect with regards to pregnancy complications?	conception_out_type_pregcomp
conception_questionnaire	What types of outcome do you collect with regards to delivery (complications)?	conception_out_type_delivery

conception_questionnaire	Do you collect data on breastfeeding?	conception_yesno
conception_questionnaire	What kind of adverse breastfeeding outcomes do you collect?	conception_brfeed_outc_pregCoh
conception_questionnaire	Please specify	
conception_questionnaire	What other types of outcome do you collect?	conception_outcome_type_others
conception_questionnaire	Could you specify the long term developmental outcomes?	conception_out_long_term_dev
conception_questionnaire	Please specify directly assessed outcomes	conception_lterm_dev_other
conception_questionnaire	Could you specify the child health outcomes?	conception_out_lterm_health
conception_questionnaire	Specify Other	
conception_questionnaire	What is the duration of follow up in your cohort for long-term outcomes? (e.g. neurodevelopmental and health related)	
conception_questionnaire	What other types of information do you collect?	conception_other_information
conception_questionnaire	Coding and Classification	
conception_questionnaire	Coding and classification of Spontaneous Data	
conception_questionnaire	Do you use specific regulations/terminology/definitions for the outcomes you collect?	conception_yesno
conception_questionnaire	Please specify	
conception_questionnaire	Do you use a coding system for classification of outcomes?	conception_class_code
conception_questionnaire	If other: what system do you use?	
conception_questionnaire	Please specify how you monitor the quality and completeness of the data	
conception_questionnaire	Coding and classification Pregnancy Cohort Data	
conception_questionnaire	Do you use specific regulations/terminology/definitions for the outcomes you collect?	conception_yesno
conception_questionnaire	Please specify	
conception_questionnaire	Do you use a coding system for drugs?	conception_code_drug
conception_questionnaire	Please specify	
conception_questionnaire	Do you use a coding system for the indication for use?	conception_code_indication

conception_questionnaire	please specify	
conception_questionnaire	Do you use a coding system for outcomes?	conception_class_code
conception_questionnaire	Please specify	
conception_questionnaire	Do you use a coding/classification system for congenital malformations?	conception_CA_coding
conception_questionnaire	Please specify	
conception_questionnaire	Quality Data	
conception_questionnaire	Screening and data analysis of Spontaneous Data	
conception_questionnaire	Do you screen the dataset periodically?	
conception_questionnaire	Which signal detection methods do you apply for the evaluation of exposure during pregnancy or lactation?	
conception_questionnaire	Which signal detection methods do you apply for the evaluation of long term outcomes?	
conception_questionnaire	Are pregnancy cases readily classified and readily to be extracted?	conception_yesno
conception_questionnaire	Please explain criteria to extract pregnancy cases from your database	
conception_questionnaire	Are exposure during lactation cases readily classified and readily to be extracted?	conception_yesno
conception_questionnaire	Please explain procedure and criteria in use	
conception_questionnaire	Screening and analysis Pregnancy Cohort Data	
conception_questionnaire	Do you screen the dataset periodically?	
conception_questionnaire	Which signal detection methods do you apply?	
conception_questionnaire	How often do you update the database?	
conception_questionnaire	Do you track Quality of the data? Please specify how	
conception_questionnaire	Do you monitor missing values?	
conception_questionnaire	Are Breastfeeding cases readily classified and readily to be extracted?	conception_yesno
conception_questionnaire	Please explain procedure and criteria in use	
conception_questionnaire	Linkage to other datasources	

conception_questionnaire	Screening and data analysis of Spontaneous Data	
conception_questionnaire	Are spontaneous reports on either exposure during pregnancy or during lactation submitted to EMA (EudraVigilance)?	conception_yesno
conception_questionnaire	Are your spontaneous reports exported to Vigibase?	conception_yesno
conception_questionnaire	Are your spontaneous reports exported to a National Pharmacovigilance Centre?	conception_yesno
conception_questionnaire	In case of export to other institutions, please specify	
conception_questionnaire	Linking with other data sources with Pregnancy Cohort Data	
conception_questionnaire	Is it possible to link your cases with other data sources?	conception_linkage_possible
conception_questionnaire	Please specify	
conception_questionnaire	If yes, to which other data sources is linkage possible?	
conception_questionnaire	Usage of the data	
conception_questionnaire	Use by third parties of Spontaneous Data	
conception_questionnaire	Are data made available for third parties?	conception_yesno
conception_questionnaire	If yes, under what conditions do you make these data available?	
conception_questionnaire	Where can applications forms for data access be found?	
conception_questionnaire	How long does the process of getting permission to use the data generally take?	
conception_questionnaire	Are there any instructions for using this data source?	
conception_questionnaire	Use by third parties of Pregnancy Cohort Data	
conception_questionnaire	Are data made available for third parties?	conception_yesno
conception_questionnaire	How do you give permission for access?	
conception_questionnaire	What is the financial policy for the use of data by third parties?	
conception_questionnaire	How long does process of getting permission to use the data generally take?	
conception_questionnaire	How long does access take after	

	permission is granted?	
conception_questionnaire	Are there any instructions for using this data source?	
conception_questionnaire	Where can applications forms be found	
conception_questionnaire	Publications	
conception_questionnaire	Key Publications with these data	
conception_questionnaire	If possible, please provide some key publications using the data (max 10)	
conception_questionnaire	Key publications of studies with these Pregnancy Cohort Data	
conception_questionnaire	If possible, please provide some key publications using the data (max 10)	
conception_questionnaire	Technical information	
conception_questionnaire	Database for Spontaneous Report Data	
conception_questionnaire	What is the database structure?	conception_structure
conception_questionnaire	Please specify	
conception_questionnaire	How often do you import the data?	
conception_questionnaire	How long are data being preserved?	
conception_questionnaire	Do you keep updating the coding systems to the latest versions?	
conception_questionnaire	In case the data model is publicly available, please provide URL	
conception_questionnaire	Database for Pregnancy Cohort Data	
conception_questionnaire	How does data entry happen?	
conception_questionnaire	What systems are in place for data entry?	
conception_questionnaire	How long are data being preserved?	
conception_questionnaire	How is security managed?	
conception_questionnaire	Do you keep updating the coding systems to the latest versions?	
conception_questionnaire	What is the database structure?	conception_structure
conception_questionnaire	Please specify	
conception_questionnaire	In case the data model is publicly available, please provide URL	
conception_organisation_type	id	

conception_organisation_type	Label	
conception_data_use	id	
conception_collection_type	id	
conception_population	id	
conception_reporter_type	id	
conception_yesno	id	
conception_data_provider	id	
conception_nature_data	id	
conception_suspect_spont	id	
conception_exposure_spont	id	
conception_coding_spont	id	
conception_adv_outcome	id	
conception_death_spont	id	
conception_preg_compl_spont	id	
conception_delivery_compl_spont	id	
conception_breastfeed_outc	id	
conception_other_information	id	
conception_longterm_outcomes	id	
conception_outcome_source	id	
conception_restr_incl	id	
conception_follow_up	id	

conception_add_info_inquiry	id	
conception_comparison_group	id	
conception_exposure_source	id	
conception_info_exposure_drug	id	
conception_info_exposure_type	id	
conception_out_type_pregoutc	id	
conception_out_type_pregcomp	id	
conception_out_type_delivery	id	
conception_brfeed_outc_pregCoh	id	
conception_outcome_type_others	id	
conception_out_long_term_dev	id	
conception_lterm_dev_other	id	
conception_out_lterm_health	id	
conception_class_code	id	
conception_code_drug	id	
conception_code_indication	id	
conception_CA_coding	id	
conception_linkage_possible	id	
conception_structure	id	

Appendix 5: Catalogue specification to register variable information and harmonisations thereof for WP7

This section describes the data model used to describe metadata on the data items used in the studies. Purpose is to track how core variables used in research are derived from source variables in the data source using a harmonisation procedure (including explanation, scientific explanation and syntax).

We will reuse the catalogue system developed by EU LifeCycle project that have used this model in practice. This project is cohort centric, a term that might not be appropriate for ConcePTION.

The model consists of:

- Cohort (describes the data source, will be merged with 'collection' metadata)
- CoreVariables (each row provides metadata on one core variable, e.g. name, type, unit)
- SourceVariable (idem, but then for data items in the data source)
- Harmonisation (describes how data is mapped from Source Variables to Core Variable).
- CodeBooks (describes the coded lookup lists for categorical variables, references from the coreVariables or SourceVariables where appropriate)

Entity	Attribute	refEntity	dataType
CodeBooks	ID		string
CodeBooks	Codebook label		string
CodeBooks	Codebook link		hyperlink
Cohorts	ID		string
Cohorts	Label		string
Cohorts	Cohort name		string
Cohorts	Basic description		html
Cohorts	Codebooks	CodeBooks	mref
Cohorts	Cohort profile		compound
Cohorts	Cohort website		hyperlink
Cohorts	Additional profile information		html
Cohorts	Contacts		compound
Cohorts	Coordinator		compound
Cohorts	Name		string
Cohorts	E-mail		email

Cohorts	Access		compound
Cohorts	Label of access URL		string
Cohorts	Access URL		hyperlink
CoreVariables	Variable		string
CoreVariables	Label		string
CoreVariables	Data type	DataTypes	categorical
CoreVariables	Values		text
CoreVariables	Unit	Units	categorical
CoreVariables	Match	Status	categorical
CoreVariables	Comments		text
CoreVariables	Harmonized variables	Harmonizations	one_to_many
Harmonizations	Harmonized variable	CoreVariables	xref
Harmonizations	Cohort	Cohorts	xref
Harmonizations	Match	Status	categorical
Harmonizations	Source variables	SourceVariables	mref
Harmonizations	Description		text
Harmonizations	Syntax		text
Harmonizations	Information		text
Harmonizations	ID		string
Harmonizations	Harmonized variable		string
Harmonizations	Source variables		text
Harmonizations	Date of update		date
SourceVariables	id		string
SourceVariables	Cohort	Cohorts	xref
SourceVariables	Label		string
SourceVariables	Variable		string
SourceVariables	Data type	DataTypes	categorical
SourceVariables	Values		text

SourceVariables	Unit	Units	categorical
SourceVariables	Collection type		string
SourceVariables	Description		text
SourceVariables	Date of update		date
SourceVariables	Dependencies		text
Terms	ID		string
Terms	Label		string