IMI2 821520 - ConcePTION

ConcePTION

WP7 – 7.1 Defining the rules and collaboration models for data reuse

D7.3 Report on data access models and rules

Lead contributor	UMCU – Marieke Hollestelle
Other contributors	UMCU – Rieke van der Graaf UMCU – Hans van Delden UMCU – Miriam Sturkenboom

Document History

Version	Date	Description
V1.0	01 Feb 2022	First draft
V.2.0	14 April 2022	Draft
V 2.2	6 June 2022	Draft
V 2.3	28 June 2022	First draft
V 3.0	15 July 2022	Final draft

Content

Document History 1

Content 2

Abstract 3

Introduction 3

Overview of research methods 5

Results 6
1.1 Literature 6
1.2 Ethico-legal documents 8
1.3 Interviews 17

General Discussion 24

Governance principles 27

References 29

Appendix A 31

Abstract

For this deliverable we aimed to better understand the general governance challenges and important governance principles for a learning healthcare system, like ConcePTION aims to develop. Understanding the governance challenges, the ethico-legal context in which data access providers work and understanding what their interests are, will be a first step for the development of a governance structure for the ConcePTION ecosystem.

To be able to provide a first step in the development of a governance structure for the ConcePTION ecosystem, we conducted a literature review study, an exploratory case study, and an interview study. We critically reflect upon the results of these studies and share important governance principles that should be taken into account.

This report presents some challenges and guiding principles for the development of a governance structure for the ConcePTION ecosystem. Hopefully, through discussions and experience a governance model will be developed and will get more refined in the future.

Introduction

ConcePTION's ultimate goal is to develop a sustainable ecosystem that can efficiently, systematically and in an ethically responsible manner, generate and disseminate reliable evidence-based information regarding the effects of medications used during pregnancy and breastfeeding to women and their healthcare providers (HCPs). This will be achieved by generating, cataloguing, linking, collecting, and analyzing data from pharmacovigilance, modelling, routine healthcare, pregnant women, and their children through a large network (ConcePTION, 2020).

The approach of ConcePTION to collect data on the safety of medicines during pregnancy and breastfeeding is similar to what is increasingly being called a Learning Healthcare System (LHS) in the literature. The Institute of Medicine defined a Learning Healthcare System (LHS) as a system in which, "science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the delivery process and new knowledge captured as an integral by-product of the delivery experience" (Olsen et al., 2007). According to the learning healthcare system movement (learninghealth.org) harnessing the power of both rigorously scrutinized guidelines with real world experience of millions of patients puts unprecedented knowledge at our fingertips – when it's needed and where it's needed, and in the form most useful to the person using it. Data, analytics, and expertise are required ingredients of an LHS. Analytics range from rigorous epidemiological studies to unsupervised machine learning, and the data science and artificial intelligence fields are booming and promising to be an advantage to medical practice for personalized treatment, value-based care, improving diagnoses and predictions.

At the same time, these advances also bring significant data governance challenges for realizing and ensuring value for all relevant stakeholders as well as individual privacy, transparency, and the protection of the public good (Winter & Davidson, 2019). Furthermore, research consortia like ConcePTION are often complex because of the collaboration between different type of stakeholders from both public organizations and the private industry. These so-called public-private partnerships (PPPs) bring together academic centers with commercial companies, often with the aim of creating large, communal resources of materials and data (Lim in Morrison et al., 2020). Companies often already have their own governance structures, reward systems and goals for extracting value, which may be difficult to align with academic goals (Morrison et al, 2020). Defining a clear governance framework in which both private and academic goals can be aligned requires the mapping of the objective and subjective interests of both type of stakeholders. Especially, when consortia like ConcePTION want to continue to exist beyond the duration of the original project and transform into a sustainable network or ecosystem. Important to consider for such a governance framework is the distributed method ConcePTION aims to use and the relationships between ConcePTION (as an

ecosystem), the data stewards (also called Data Access Providers (DAPs)), and the pregnant and breastfeeding people, whose data is of great importance for the ConcePTION ecosystem.

For this deliverable we aimed to better understand, the general governance challenges and important governance principles for research collaborations like ConcePTION by creating a brief overview based on the literature. Furthermore, we aimed to understand what kind of *ethico-legal* documents DAPs have, to better understand already existing organizational governance procedures. Lastly, we aimed to get an understanding of the objective and subjective interests of the DAPs and partner organizations of the ConcePTION ecosystem and explore their views on their role within the ConcePTION ecosystem. Understanding the governance challenges and the interests of participating organizations will help with the development of a governance framework for the ConcePTION ecosystem which can hopefully support the collaborative initiatives, the return of results to pregnant and breastfeeding people and their HCPs as well as the improvement of the clinical practice. Having a governance structure can be a way of ensuring that there are clear rules about data access, publication, patenting, and how other activities will operate; and it can delineate responsibilities (Morrison, et al., 2020).

How to read this document

This document presents an overview of the work for task 7.1 in Workpackage 7 of the ConcePTION project and provides guiding principles and norms for activities within the ConcePTION ecosystem from a governance perspective. These insights could be used to further develop a governance framework for the ConcePTION ecosystem. Besides the guiding principles and norms presented in this report, we need to take into account that researchers and data access providers should handle data in accordance with many varying conditions, like conditions which originate from local laws and regulations, data governance arrangements, informed consent procedures, and research procedures and protocols. In addition, we also need to take into account that there can be potential changes over time in technology, resources, data protection legislation and interpretations of the GDPR. Lastly, we hope that through discussions and experience a governance model will be developed and will get more refined in the future.

Governance: a definition

The term governance is used in many shapes, in many disciplines, and it pops up in a lot of literature on how to create responsible big data collaborations or structures. In the latter, governance is often referred to as missing, or at least, it has been mentioned as essential for the development of a sustainable and ethically responsible PPPs, Big data projects in health care, learning healthcare systems and the like.

The question rises, what does governance mean and is there a definition for the report? Again, in the literature, one can find many interpretations and formulations of the term governance. From the literature that include definitions of ‘governance’, as well as the literature on what good governance could entail, an understanding of governance was developed by colleagues of the UMCU, department Medical Humanities. According to them, governance consists of organizing interaction and decision-making processes, as well as delineating responsibilities and tasks to facilitate appropriate conduct, oversight, and cooperation. In this report, we focus mainly on the tasks to facilitate appropriate conduct, oversight, and cooperation.

ConcePTION

Throughout this document, ConcePTION refers to the current group of experts that work towards the goals as formulated in the description of action of the ConcePTION project. We talk about the ConcePTION project, which refers to the goal of ConcePTION and the work that needs to be done to

reach the goal. We talk about the ConcePTION network, referring to all the partner organizations and third-party organizations and other cooperative organizations within the ConcePTION project. Lastly, we refer to the ConcePTION Ecosystem and ConcePTION as an LHS, which will be the entity that hopefully will keep existing after the project time has ended.

This document aims to facilitate and encourage further deliberation but does not by itself guarantee ethically appropriate data-intensive research practices. All stakeholders remain responsible for compliance with national and international laws, general ethical principles, and best practice standards that respect restrictions on downstream uses.

Overview of research methods

To be able to create a first step towards the development of a governance framework for the ConcePTION ecosystem, we conducted a literature review study, an exploratory case study, and an interview study. We will critically reflect upon the results of these studies and determine what important governance principles are, based on our knowledge on the goal and methods used within the ConcePTION project, as well as learning healthcare system literature.

For the literature search, we took a pragmatic search strategy to identify key themes that could be explored further in the case study and interviews. We searched literature on the governance of LHSs, data intensive research, public-private partnerships, responsible data sharing, and the ethics of LHSs.

For the exploratory case study, our aim was to understand pre-specified governance procedures captured within ethico-legal documents. For this study, we followed a three-step approach. First, we identified the DAPs in ConcePTION, who have data or have access to relevant data for the ConcePTION ecosystem. We then attempted to obtain key ethico-legal documents. Ethico-legal documents are documents that contain some statement on the conditions for data collecting, processing, and sharing, can be shared with the researcher. Lastly, we categorized the documents and used the Council for International Organisations of Medical Services (CIOMS) guideline 12, to explore what ethical principles are already upheld within these organizations and what could be improved (CIOMS, 2016).

For the interview study, we conducted semi-structured qualitative interviews with a range of stakeholders who own and/ or have access to relevant data within the ConcePTION project. We aimed to include people who are working as partnering organizations and third parties in the ConcePTION project. To be able to invite different DAPs, we distinguished between: private (pharmaceutical company and private centers) and public organizations (universities, teratology information centers, public health services, and hospitals), countries, regions, collaborative partnerships, and occupations.

In deliverable 7.23, the concept of an LHS has been thoroughly discussed. In an LHS, care and research are aligned to accelerate research and outcomes for patients and to overcome current problems, such as low inclusion rates and complex informed consent procedures. LHSs aim to systematically study, evaluate and improve quality and efficiency of care while speeding up the process of generating generalizable medical evidence for pregnant and breastfeeding people. Furthermore, in deliverable 7.23, we discuss an ethics framework for an LHS for pregnant and breastfeeding people, which contains important ethical principles to help guide the development of an LHS for pregnant and breastfeeding people, which in turn could be used for the development of the ConcePTION ecosystem. For the development of the governance framework, the ethics framework will also be consulted.

Results

1.1 Literature

Method

We have consulted the literature to explore important governance issues for LHSs. Since ConcePTION is a public-private partnership, which aims to execute data intensive research by means of a federated system, we used literature discussing many of these different characteristics to inform important governance challenges.

Results

In the literature a lot has been written about challenges, concerns, expectations and recommendations for LHSs, or (health)data-intensive research initiatives. From the literature, we can summarize a few important principles that are relevant for a governance framework for ConcePTION as an LHS.

Table 1 overview principles and literature

Ethical oversight	Wouters et al., 2021; McLennan et al., 2018; Piasecki & Dranseika, 2019;
Transparency	Jones et al., 2020; Rossi & Lenzini, 2020; Botkin, 2017; Morrison et al., 2020; Spagnuelo et al., 2019; Hollestelle et al., 2022; Torchia et al., 2015; Ballantyne & Stewart, 2019; Xafis et al., 2019;
Privacy and confidentiality	Kuchinke et al., 2016; Wouters et al., 2021; Rieke et al., 2020; Mann et al., 2016; Kalkman et al., 2019
Informed consent	Spectator-Bagdady & Jagsi, 2018; Wouters et al., 2021; Piasecki & Dranseika, 2019; Cheah & Piasecki, 2020;
Social value/ social license	Kalkman et al., 2019; Moloney et al.., 2016; Muller et al., 2021
Accountability / liability / responsibility	Wouters et al., 2021; Rieke et al., 2020; Morrison et al., 2020;
Stakeholder engagement	Jones et al., 2020; Morrison et a., 2020; Seid et al., 2014; Torchia et al., 2015; Ballantyne & Stewart, 2019;
Trust	Kalkman et al., 2019; Botkin, 2017; Hollestelle et al., 2022
Return of results	McLennan et al., 2018; Hollestelle et al., 2022
Legacy planning and sustainability	Morrison et al., 2020

Learning Healthcare System

In the literature on the governance of LHSs, attention is often brought to the challenges, concerns,

benefits, and expectations of implementing an LHS structure (Jones et al., 2021; Wouters et al., 2021; Botkin, 2018; Piasecki & Dranseika, 2019; Wouters et al., 2020;). It is to be expected that the implementation of an LHS would advance medical knowledge and would improve treatments, quality of care and the availability of care. From different qualitative research as well as from our own interview study with women during preconception, pregnancy, and nursing, it became clear that most patients believe an LHS to be most useful for healthcare professionals, so that they can receive better knowledge to base health information and recommendations on (Hollestelle et al., 2022; Jones et al., 2021).

Despite the promises to improve clinical care and accelerate scientific research, transforming the healthcare system into an LHS raises challenges and concerns. Challenges range from methodological challenges such as the quality and curation of the data and data analyses (because of, for example, confounding variable, missing data and observational errors) for improving care and the individual translation of results, to ethical challenges such as the evaluation of an LHS, transparency, informed consent procedures, and the support of a broad range of key stakeholders in the health systems (Wouters et al., 2021; Brody & Miller, 2003; Moloney et al., 2016). With that, most patients are concerned with the possibility that profit-driven users of LHS data become too involved in ways that could burden or exploit patients, hinder medical objectivity, or compromise patient-physician relationships (Hollestelle et al., 2022; Kalkman et al., 2019).

Public-private partnerships

One of the main characteristics of ConcePTION, and possibly one of the interesting new characteristics for an LHS, is that the ConcePTION consortium is a public-private partnership (PPP). In the literature, there are many different descriptions used, all describing the often-structured collaboration between public and private actors for the achievement of common goals (Klijn & Teisman, 2003). Important, there are three main features, relevant to the ConcePTION PPP: 1) collaborating actors within PPPs, have different ownership structures: namely some are public organizations, and some are privately owned organizations. 2) outcomes from PPPs are always public or quasi-public goods and services for the benefit of a third party. In this case: for the benefit of different stakeholders, such as pregnant and breastfeeding people, their HCPs and regulators, and 3) a part of the PPP remains in effect for a longer period of time after the project has ended (Torchia et al., 2015). In the last decade, the use of PPPs has been increasingly popular, also as established components of the health system (Ballantyne & Stewart, 2019), but also especially for tackling problems that are too complicated for individual actors or sectors. A review conducted in 2013 by Torchia and colleagues, explored PPPs and which main drivers and characteristics lead to the success of PPPs. A few important insights from the review are:

a) That the **effectiveness** of PPPs is positively influenced by: an active role of governments in regulating the sector in PPP projects in health care, transparency and sound regulatory frameworks regarding private sector participation, emphasis on stakeholder involvement and evaluation on effectiveness before implementation and to review crucial elements such as: costs (costs in PPPs are frequently underestimated), quality (quality is almost always the trade-off in comparison with costs and time, since they are oftentimes fixed in PPP contracts), flexibility (health care changes rapidly, while PPP contracts or rigid) and complexity (agreement among all different types of stakeholders within PPP is extremely difficult, perhaps these models can be simplified to overcome this complexity).
b) That while PPPs combine the best of both worlds and can produce innovative methods in health care, consideration needs to be given to whether the PPP will deliver the desired **benefit**.
c) That **public interest** is one of the main concerns relating to PPPs in the health care sector. Because public interest may not necessarily be the primary goal of the private sector, there is a need to develop a set of global norms and ethical principles. With that, the driving principles of the partnership should be rooted in the *benefit for the society* rather than the *mutual benefit for the partners.*

Another paper by Ballantyne & Stewart discusses three specific challenges for PPPs in Big Data,

namely: 1) working within the social licence (especially challenging because the social licence of the public sector data use may not extend to private sector use), 2) public antipathy to the commercialization of public sector health data, and 3) questions of ownership, both of the data and any resulting intellectual property or product (Ballantyne & Stewart, 2019). Furthermore, they have used the Deliberative Framework, to identify ethical values that are engaged by PPPs and which could help navigate through tensions between values (Ballantyne & Stewart, 2019). Besides the three specific challenges mentioned earlier, the framework also discusses procedural values, such as transparency, accountability, trustworthiness that underline the tension with the use of (Big) data in PPPs (Xafis et al., 2019; Ballantyne & Stewart, 2019).

Responsible data sharing

Creating an LHS needs to be done in a responsible manner. Part of this, is creating a good governance for data research activities. Although ConcePTION uses a federated system for data analyses, and therefore, individual level data will stay local and will not be shared among partnering organizations, it is useful to explore literature on principles and norms for responsible data sharing in international health research. Kalkman et al (2019) systematically reviewed principles and norms for responsible data sharing in international health research. Four overarching themes were noted: societal benefits and value, distribution of risks, benefits and burdens, respect for individuals and groups and public trust and engagement. Furthermore, the themes are followed by principles and the respective norms promoting these principles. However, the article by Kalkman and colleagues also raises the questions how to harmonize the identified principles and norms into a coherent governance framework. For the creation of a good governance structure within ConcePTION, these overarching themes are a good starting point.

To have a better understanding of the context in which partnering organization already work, we collected so-called ethico-legal documents, which will be described in the next section.

1.2 Ethico-legal documents

For the development of a governance framework, it is important to understand the context of the participating organizations in terms of ethical, privacy and governance rules they need to adhere to before any type of collaboration, research activity or data sharing. Therefore, we asked for the so-called ethico-legal documents, to gain insight into pre-specified conditions for data sharing, site-specific legal and ethical conditions for data sharing, and additional efforts for the development of a governance framework for international data sharing.

As mentioned earlier, the literature describes different challenges and concerns regarding the implementation of an LHS, and more fundamentally, working within such an LHS regards clear rules for data collection, access and potentially data sharing.

Method

For this study we attempted to obtain any relevant material in terms of ethico-legal documents from organizations who own or who might have access to relevant health data (also called Data Access Providers (DAPs)). With that, we send a checklist (see appendix A) with common standard documents that would fall under the scope of ethico-legal documents, which organizations could use to either state the existence of such documentation, describe their documentation, and share their documentation.

All documents that include a statement on the conditions for data collecting, processing, and sharing, could be send to the researcher. Examples of those documents are:

Policy document(s):

1. **Policy statements,** including national and international statements (e.g. on the implementation of the GDPR in your organization and how personal data is processed). Also known as: *privacy statement, Data protection policy, privacy policy*.
2. **Governance structure for each of the data source(s) that your organization will access for ConcePTION** (e.g. who is deciding on access who holds the data: e.g. Data privacy officer, Data controller, Data protection officer, team of researchers)

Standard Operation Procedures (SOPs):

3. **Service agreements** to get access to data or share data inter- or intra institutional (e.g. Data processing, Data access and transfer agreements (DPAs, DAAs and DTAs))

Research-related SOPs:

4. **A description of the requirements of local, organizational and/ or national governance, ethics, scientific or regulatory committees** that need to approve study protocols prior to protocol execution in the data sources you will use/access for ConcePTION
5. **Standard template protocols** that above mentioned committees need to approve before conducting the research
6. **Study protocols** that are written before conducting research that follows the outline of the document mentioned above
7. **Patient information forms** that are used by those collecting the data to inform patients that their data is stored and used for research
8. **Informed consent forms** that patients may need to sign before their data is stored.

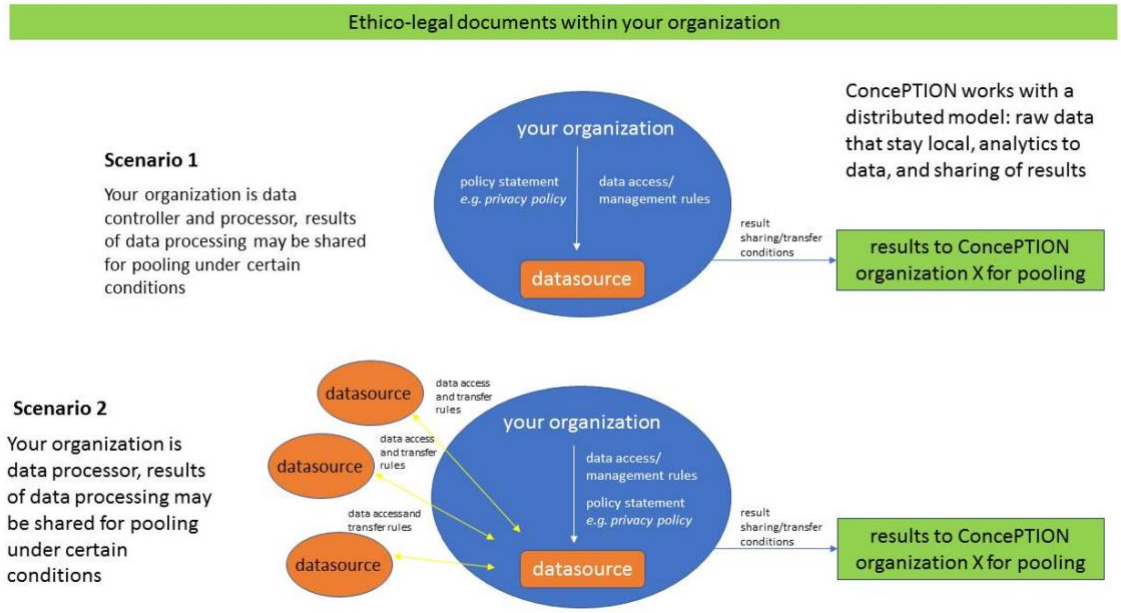

Figure 1 Ethico-legal documents scenarios

This figure shows two scenarios, of which at least one of the two applies to an organization. Based on these scenarios, it will become clear what kind of documents the organization has.
There are (1) **policy documents** within your organization, stating for example: how data is collected, used and protected, (2) **standard operation procedures** (SOPs) concerning data management, and service agreements within your organization or between organizations (data sources) that state the terms and conditions for data access and data sharing between sources and information on the type

and characteristics of these data, and (3) **research-related SOPs**, (when research is conducted within your organization) these documents communicate how new data is collected, stored and used.

After receiving the ethico-legal documents and the checklists, we categorized the shared documents to make an assessment and overview and analyzed the checklists and content of the shared documents, with a particular focus on overlap and divergence.

Furthermore, we extracted information from the documents and used multiple viewpoints for analyzing the documents.

First, we compared the different documents and checklists form the DAPs and counted how many documents we received, and more importantly, what type of documents we received. Second, we used the CIOMS guideline 12 on the collection, storage, and use of data in health-related research, as a framework for analyzing the ethico-legal documents and we extracted information along the lines of the following key elements: statements on confidentiality and level of de-identification, purpose limitation, statements on data sharing and data access, and governance/ review procedures.

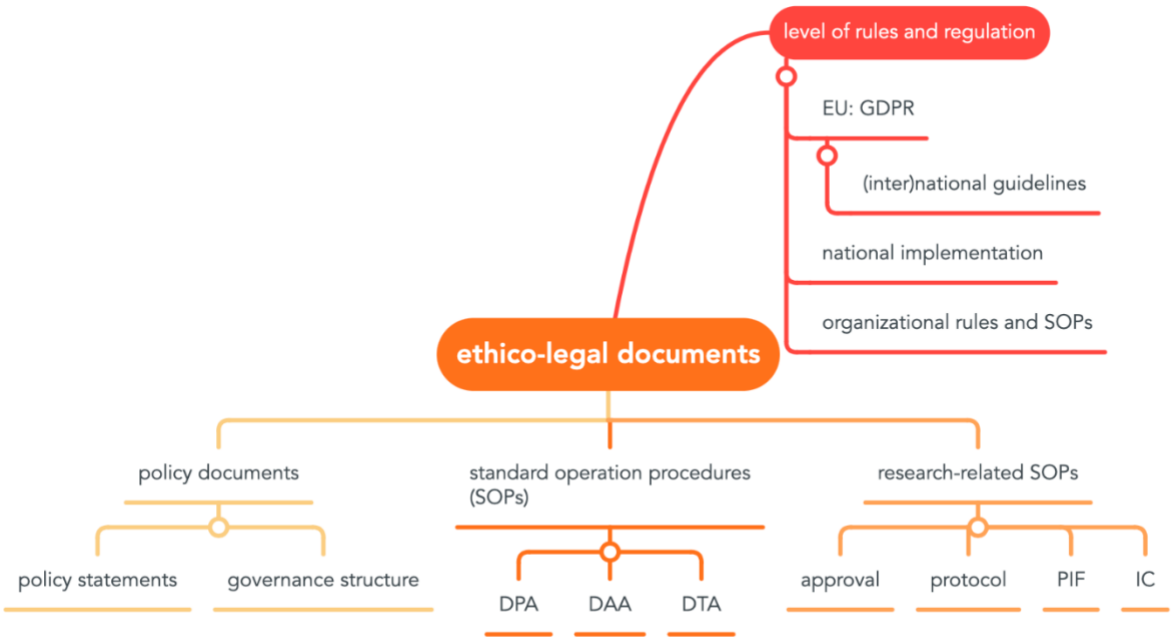

Figure 2 Overview of ethico-legal documents

Retrieval of the documents

All 18 DAPs who were participating in WP7 were contacted via the ConcePTION Task Management System (TMS) and asked to send their documents. In cases where documents had been provided, we noted the number and type of documents that had been shared with us.

Results

Documents

Out of the 18 DAPs that were contacted via the TMS, 15 DAPs responded and shared some documentations plus the checklist. We received 65 documents and some references to webpages containing relevant information. After reviewing and organizing the documents the received documents can be categorized as shown in Table 2. Some documents overlapped in category, meaning that a policy reference could have been part of a governance description. Furthermore, from the filled in checklist, we can show what DAPs stated to have in Table 1.

Some DAPs stated to have a certain type of documentation, but decided not to share it with us because of the following reasons: documents were written in a different language (other than English, German or Dutch), it involved undocumented procedures, documents are available elsewhere (for example on a website or with another institution), activities are reviewed internally, some documents could not be shared, and some did not explain why certain documents were not shared.

Table 2 Received Ethico-legal documents

Documents*	
Questionnaires/ registration forms	4
Patient Information Form	2
Informed Consent form	3
DAA/DPA/DTA	22
Governance	6
Policy references	31
Research SOP	3
Other	4
Other/ policy	14
Documents/ websites	65

Table 3 Ethico-legal documents DAPs have stated to have

What kind of document?	***Yes***	***No***
Policy documents	*12*	*3*
Governance structure for data sources	*8*	*7*
DPA	*10*	*5*
DAA	*8*	*7*
DTA	*7*	*8*
Description of approval committees	*9*	*6*

Standard template protocol	*7*	*8*
Study protocol	*6*	*9*
Patient Information Form	*3*	*12*
Informed Consent form	*2*	*13*
Other	*6*	*9*

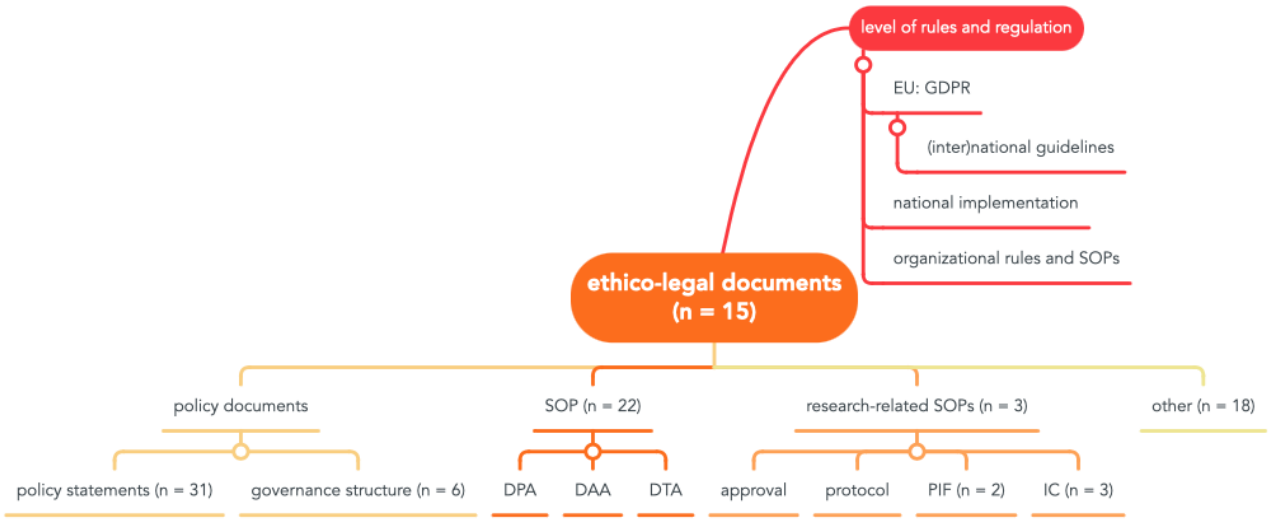

Figure 3 Overview ethico-legal documents

Analysis of documents

For the analysis of the documents, we have chosen the CIOMS guideline 12 as a framework. In the COIMS guidelines, ethical principles are set forth that should be upheld in the ethical review of research protocols. The guideline also recognizes the value of data collection and formulates items that need to be regulated by institutions where data is collected and/ or archived. We, therefore, wonder whether DAPs have an ethico-legal structure that regulate these specific items.

Statements on confidentiality and level of de-identification

Most organizations state something regarding the confidentiality of the link between collected data and personal identifiers of the donor and explain the procedure for maintaining confidentiality. Most documents stated to only use anonymous data or that data is anonymized before sharing it with anybody. Other documents explicitly mention “pseudonymized”, “encrypted”, “de-identified” or “aggregated” data as the type of data that can be used for the research within projects like the ConcePTION project. Some documents that stated to only use anonymized or anonymous data used the following concepts for anonymous data: “i.e. data at an individual personal or corporate level”, “i.e., data for which all identification details such as name, ID number and address have been removed.”, “that cannot be traced directly or indirectly to a person and/or healthcare

providers/institutions.", "not containing variables with personal data". With that, only a few documents refer to national laws that prohibit the use of any other type of data than anonymized data.

Another document explicitly explained how identifiable data is processed and linked to other data. In this case, a non-profit foundation manages and processes the patient data. This foundation is also responsible for removing identifiable patient data before sharing the data for research purposes.

Few organizations also included an informed consent form, which contained an explanation for participants on how their data is being protected and safeguarded. Such a statement included technical (online security, avoiding risk of data loss, alternation of data or unauthorized access, avoiding risks associated with processing personal data) and organizational measures (restricting access to personal data solely to authorized persons with a legitimate need to know).

Another organization explained how anonymized data cannot be accessed until one year after the reference period.

Purpose limitation

A lot of documents combine the information on confidentiality with a description for which purposes data can be used or requested. Almost all documents stated that data can only be used for (scientific) research purposes. However, it was not specified for what type of research.

In other documents, such as the database description documents of the DAPs for ConcePTION, DAPs state for what type of research data can be used, and with that, for what type of research their data *can* be used.

Since we have not received many informed consent forms or patient information letters, we do not know what DAPs or other organizations that collect the raw data explain to data subjects in terms of what type of research is going to be conducted with their data. In the few patient information letters we did receive, it is oftentimes mentioned in what (research) context data is being collected. For example: in the context of collecting information on congenital anomalies.

Statements on data access and sharing

A lot of organizations include a statement on data access and data sharing. These statements explain roughly who has access to what type of data and what type of data can be shared with whom. One organization explicitly states that access to micro data can be granted to researchers and analysts from the same or pre-approved institute. Authorization can be granted to both public and private organizations. Furthermore, private organizations must have a stable research or analysis environment, which includes a responsible manager and a group of researchers.
Most documents restrict use of stored data to scientific research within the scope of primary research activities or goal only, either by limiting use by partnering organizations, already approved projects, and relevance to the dataset or original goal of data collection itself.

Some organizations state that they exclusively grant access to the data to researchers employed by universities and research institutes for scientific research purposes.

Almost all documents state that personal data are never shared with other researchers or institutes and can only be used by persons who are directly working on the specific research project, of for whom approval was granted or who works within the same institute.

One organization also included a document which includes a statement on how a researcher can apply for data access and reuse and provide for a list of questions research should be able to answer before applying for the data.

Some organizations state that the collected information will not be given to any third party unless it is required by law.

Review/ governance

A lot of organizations do have some sort of statement explaining which body reviews research proposals for future use of the data.

Organizations mention that approval is needed from: a Health Data Board/ Institute, a Data Protection Agency/ Authority, a Medical Research Ethics Committee, register administrators, or/ and compliance with national law and guidelines need to be guaranteed.

Most organizations refer to compliance with laws and regulations. Although, the GDPR only came into force as of 25 May 2018, not all organizations refer to the European regulation. The vast majority of documents make reference to national or even regional legislation. General statements often encountered is that "data will be correctly used with compliance with laws and regulations with regard to guaranteeing the privacy and confidentiality of data and the applicable national standards".

Underrepresented topics

Not all topics from the CIOMS guideline 12 were represented or only discussed by one or two DAPs within the ethico-legal documents. The reasons for why some DAPs did not share documents containing these topics could have to do with the fact that they are not data controllers themselves, or that they do not conduct research for which certain documents with information are needed. It could also be the case that DAPs have documents or other types of information sources which does include information on these topics, but because we did not encounter these topics, it is difficult to say something about the existence and the way DAPs have arranged these items.

Although most ethico-legal documents explained how confidentiality of the data is being handled, it is not always clear from the documents how data subjects are protected or how data subjects can retract permission for using their data for all types of purposes. In the CIOMS guidelines it is stated that data subjects should be able to withdraw their consent for use of data in a databank, which should be formalized by written documentation signed by the data subject or a legal representative.

Another item we did not find in the ethico-legal documents, is in which circumstances DAPs need to recontact data subjects. Recontacting data subjects is sometimes necessary when consent for a new use needs to be obtained. There are also cases for which recontacting is more complex. For example, when researchers want to use data from health-related registries, it might be that these data subjects are unaware that their data were submitted to the registry or are unfamiliar with the process by which these researchers can obtain access to the data. In principle, these researchers should ask informed consent for these studies.

None of the ethico-legal documents contained information on whether there are appropriate mechanisms for keeping data subjects informed of research outcomes, to whom any benefits from the research are expected to accrue, and whether, and if so, how participatory engagement with patient groups or the wider community is organized. One explanation for the lack of information on these items could be that it was not specified in the checklist, and therefore, DAPs did not search for documentation on these topics. Another explanation might be that these items are covered in other type of documentation. Or, like mentioned earlier, since they are not the data controller, they feel it is not their responsibility to provide information to the data subjects or arrange participatory engagement with patient groups.

In principle, informed consent forms or (patient) information leaflets, should include a statement on whether return of information from data analyses is foreseen, and whether the donor wishes to receive that information. With that, for the collection of data for the purpose of routine clinical care or health registries, the informed opt-out procedure is often used. With informed opt-out is meant that data may be stored and used for research unless an individual explicitly objects. At the same time, opt-out is sometimes not applicable when it is mandatory to include data in population-based registries. Important for informed opt-out is that: 1) people are aware of the existence of the registry, 2) sufficient information is provided on what data and how data is collected for what purposes, 3) people need to be informed that they can withdraw their data, and 4) there is a genuine possibility to object offered (CIOMS, 2016).

Concluding remarks

For this study, we received 65 documents that have some information about ethical and legal conditions of organizations who own or/ and have access to relevant data for the ConcePTION ecosystem. We studied the documents and tried to categorize them. Furthermore, we used the CIOMS guideline 12 to see whether DAPs have information on certain items that are, according to the guidelines, important items for consideration when collecting, storing, and using data in health-related research.

Interestingly, from the received documents it seems to be most important for DAPs to provide information on how confidentiality of data subjects is being protected and that access to (raw) data is oftentimes not possible, or requests need to be reviewed by a type of review committee.

Furthermore, we observed convergence on the conditions for data sharing, which is oftentimes only permitted for scientific research, in anonymized or in a coded/ pseudonymized form, after approval from a designated committee, with compliance with the laws and regulations (both regional, national, and European) that are applicable. This finding is in line with earlier research on the topic of responsible data sharing in health research, performed by Kalkman et al. (2019) for the BigData@Heart platform. Researchers also found convergence on these topics in the ethico-legal documents of the organizations working within the BigData@Heart project. Apart from this consensus, there are multiple challenges and lessons learned for the ConcePTION project.

Challenges and lessons learned

Terminology

As explained earlier, DAPs use different terminology, and the terms pseudonymized, coded, and anonymized are used interchangeably. With that, as shown above, it appeared that different descriptions are attached to the same terms. This is particularly relevant, since the GDPR for example, only applies to personal data. Under the GDPR, coded/ pseudonymized data is also considered personal data, however, anonymized data is not and does not enjoy protection form the GDPR. The question rises, whether DAPs have used the term “anonymized” correctly, since a lot of data might come from electronic health records or research databases, which probably is personal data. With that, pseudonymization is reported by some organizations as a sufficient reason to not ask informed consent for future research initiatives. This might be true, if the key to re-identification is not being shared and broad consent has been asked. Also, it seems to be the case that most documents are legal documents or focused on laws and regulation regarding the collection, use and storage of health data. Within ConcePTION it should be made clear how to deal with varying levels of de-identification and ConcePTION could encourage DAPs to consider clarifying terminology on confidentiality of their data.

Access to data

For some DAPs, access to data is not always possible, not even for scientific research purposes. Or when access to data is possible, like mentioned before, requests need to be made and reviewed by a designated committee. Some documents explicitly mention whether scientific research for commercial purposes is permitted. Although it is valuable for DAPs to have these rules written down, they are rather complex, and even form obstacles for participating in a public-private partnership. Once ConcePTION starts to generate results or when the ecosystem is established, it is important to decide on the matter of who gets access to what type of information and in what ways. If some DAPs have issues with working together with the private industry, how is ConcePTION going to make sure both interests are being protected to uphold the collaboration?

Another point relating to access, is that some purpose limitations are quite limiting. The question rises how specific the description needs to be of the type of research that is going to be performed in the ConcePTION ecosystem.

Patient/ participation information

Not a lot of DAPs send us documents that include information for patients or participants. There might be reasons for why DAPs do not have such documents, but that does not mean they have no obligation to the data subjects at all. It is important that DAPs are familiar with the processes of informed consent, return of results, and who benefits from the results of their research.

Soft law

Only a few of the policy documents included ethical principles, such as: integrity, transparency, and FAIR principles for research. These principles are not always articulated in laws and regulations, but are extremely important to protect the quality and trustworthiness of the ecosystem, but also to protect the rights and welfare of the data subjects.

Limitations

Our collection of documents is by no means exhaustive or comprehensive. We asked for ethico-legel documents, which is by itself not a distinct concept. Although we explained what types of documents we were looking for, it could have been the case that DAPs were unfamiliar with the documents or some type of them. With that, we might also have had limited understanding of what type of documents exist at the different organizations, and therefore, missed out on some document containing information on the items we did not come across in the documents as much. Nevertheless, our results do reflect the conditions as specified in a representative sample of ethico-legal documents.

1.3 Interviews

There is relevant literature on challenges for PPPs and LHSs, and also literature that reflects the debate on topics of sustainability, effectiveness, efficiency, and benefits. However, we are also interested in what the organizations that own or have access to relevant data for the generation of new knowledge, think.

Method

Design

We conducted a qualitative study design to collect the views and reflections of the data access providers (DAPs) and the partner organizations of IMI ConcePTION. We performed semi-structured interviews with a topic list (see Table 4). The topic list was based on the topic list used for another qualitative interview study within WP 7.3, in which we asked women during pre-conception, pregnancy and nursing what they think about an LHS for pregnant and breastfeeding women (Hollestelle et al., 2022), and on discussions amongst the research team.

Table 4 Topic List

1. Willingness to participate
2. Expertise and double role function
3. Future
4. Conditions for working
5. Added value

Sample and Setting

To capture a wide range of interests and perspectives (contrast maximization), a variety of people from different types of organizations and different countries were identified. We aimed to include people who are working as partnering organizations and third parties in the ConcePTION project. To be able to invite different DAPs, we made a distinguish between: private (pharmaceutical company and private centers) and public organizations (universities, teratology information centers, public health services, and hospitals), countries, regions, collaborative partnerships, and occupation. Respondents were recruited by means of purposeful sampling. Respondents were approached by e-mail. Most of the interviews started with an introductory question relating to the work of the respondent and how data collection, storage, and analysis work within their organization. We then used the topic list to continue with the interview. The interview allowed respondents to bring up or emphasize particular new issues they considered relevant. The interviews were conducted in English and Dutch and took place via Microsoft Teams.

Data analysis

After transcription, we analyzed the interviews according to the thematic analysis method and by going forth between data collection and analysis to develop codes (Charmaz, 2006). An initial coding list was developed based on the topic list. Subsequently, we coded the transcripts. The coding list was evaluated and adapted, and all interviews were coded using Nvivo 12 software. An independent researcher also read and coded 8 out of 14 pseudonymized interviews to validate the results of the interview analysis. In the course of analysis, codes were adapted, and additional codes were added to the coding list where necessary. A meaningful pattern was identified across the data set, leading to the formulation of interpretative higher order themes. The themes capture the views and interests of DAPs regarding the ConcePTION ecosystem. The themes represent both topics that were often discussed by participants and a variety of views that are of help in the development of a sustainable ecosystem of continuous learning. The findings, including the coding list and formulated higher order themes, were discussed within the complete research team.

Results

Out of the 23 DAPs that were approached, 14 agreed to participate in the study, 4 were unable and 5 did not respond. A total of 14 semi-structured interviews were conducted with 18 people involved in IMI ConcePTION. Two of the DAPs were represented by two employees of the same organization or research collaboration. Interview respondents worked in different organizations, including universities, public health centers, hospitals, teratology information centers, pharmaceutical companies, and private centers. Table 2 shows all relevant characteristics of the respondents. We could not share all details, in order to ensure the privacy of the respondents.

Table 5 Characteristics of the respondents

Interview number (I)	*Type of organization*	*Public/ private organization*
1	University	Public
2	Research institute	Public
3	Pharmacoepidemiologic research institute	Public
4	Research institute	Public
5	Hospital	Public
6	University	Public
7	University	Public
8	Pharmaceutical company	Private
9	Public Health Service	Public
10	Pharmaceutical company	Private
11	University	Public
12	Neurodevelopmental research, pediatrician	Public
13	Centre of Health	Private
14	University	Public

Three main themes relevant to this report were identified during the data analysis and they will be discussed below followed by some concluding remarks.

Theme 1: Motivation for participation

Most participants seem to contribute to ConcePTION because they view the public-private partnership as an *opportunity* to: 1) contribute to the goal of creating knowledge on the safety of medication during pregnancy and breastfeeding, 2) look at medication safety, birth defects in a bigger context (namely European wide), 3) collaborate with other registries, databases and the like, 4) stimulate scientific research, 5) to learn from others and their registries, and 6) showcase their own databases and share expertise.

Almost all participants mentioned they want(ed) to participate in the project because they want to work together on this subject and want to learn how other organizations manage their databases, so that they can either share their expertise or improve/ professionalize their own databases. Working together was seen as beneficial for the goal of knowledge generation and as something positive and exciting for their own personal motivation and working experience.

Participants emphasized that current initiatives are stuck and do not advance enough, and that there is now the opportunity to learn from real life data. It was mentioned by some participants that they feel it is their responsibility, or ethical obligation to contribute, because of the database or resources they have. They felt that they, with their organization, are in the position to contribute to something important, and therefore they must. All participants felt some responsibility for the group of pregnant and breastfeeding women and their offspring, but not all mentioned this responsibility towards pregnant and breastfeeding people as a motivation to work or keep working for ConcePTION.

Regarding the sense of responsibility, there was also a difference in the articulation of that responsibility between public and private organizations. Besides the responsibility for pregnant and breastfeeding women and their offspring, the participants of the private industry also explained that they need to generate knowledge, because it is a requirement from the EMA and FDA. Because they are required to research medication safety among pregnant women, this was considered to be another type of obligation and with that, a different type of motivation.

Theme 2: Conditions for participation
During the interviews, we talked about the conditions the DAPs have to keep working for the ConcePTION ecosystem. This theme can be split into the most discussed conditions.

Resources
In all interviews, financial resources were discussed as a condition for participating in the ConcePTION ecosystem. Interestingly, besides the obvious (the need for financial resources to actually work for the project), financial resources were mentioned as important for many different reasons. Financial resources were discussed in the following ways:
(1) As a stable flow of income. Preferably agreed upon with a contract for a longer period, covering all the planned activates. A stable flow of income is both beneficial for attracting and training more employees in this area of work in order to distribute tasks and to become more specialized in the field of pharmacoepidemiology. It is also necessary for planning ahead and being less dependent from other financial sources to keep “the system running”, which oftentimes takes up a lot of time that cannot be used differently. Most importantly, a stable flow of income also means that they can keep working for the ConcePTION ecosystem and do not have to commit to other organizations, consortiums, tendering calls etc.

(2) In terms of funding and different ways of funding. Some organizations state specifically that they cannot receive funding from the private industry. These organizations are, for example, independent (public) institutes who state that there would be a conflict of interest. With that, there are different types of organizations that rely on financial resources in a different way. Some DAPs mentioned that in terms of funding from the pharmaceutical industry, it would be best if an independent party would review the funding and would, among other things, make sure not one specific center would get paid for a study.

Other participants mentioned that besides financial resources, they would also need IT/ computational resources for doing the actual analyses and for making sure they can keep up with the heavy computational work which is necessary for sustaining the Common Data Model (CDM).

Scientific Input and motivation
Another condition mentioned by some participants is the importance of scientific input and the ability to publish. Some participants are academic institutions, which depend on producing scientific publications. Therefore, if they would keep working in the ConcePTION ecosystem, they would have to be able to also publish new insights resulting from the ecosystem in scientific journals.

With that, some participants also emphasized the need for posing more scientific questions and implementing more scientific methods within the ecosystem. Working for the ConcePTION ecosystem should be different from tendering for pharmaceutical companies. Most participants argued that it is much more gratifying to come up with a scientific question and method and that this is something that motivates them much more.
Motivation was also mentioned by a few participants as a condition for working for the ConcePTION ecosystem. Participants want to feel motivated in order to keep working for the ConcePTION ecosystem. Motivation, according to some, is stimulated in different ways, but most importantly by: (scientific) interest, autonomy, trust, and good working relations.

Support
Some participants mentioned they are resource constrained given that some DAPs are not used to

writing certain types of protocols or experience challenges with receiving ethics approval for studies.

It was also mentioned by some participants that they would like the ConcePTION ecosystem to have a permanent staff for these kinds of support and to be able to ask questions regarding timeline, deadlines, funding, ethics, and events.

Personal relationships
Participants also mentioned having personal relationships with other people working for the ConcePTION ecosystem as an important condition. Most of them argued that it is nice to know whom you work with and who is giving new deadlines for certain tasks. Knowing the other person, a little bit more than just via online communication, will improve the communication, but also has a positive effect on the willingness to do the work. Furthermore, other people have the ability to inspire you and with that, have a positive effect on how fast you are willing and able to do and complete something.

Almost all participants mentioned that up to now, the pandemic has created a huge barrier for meeting people, and therefore, for creating some sort of connection with other people working for the ConcePTION project. Many people miss the personal interaction which oftentimes is stimulated with conferences and with live meetings. Right now, it is too easy to just turn off your camera, join a meeting without contributing. Therefore, participants also argued that it is important to them that people show that they are actively involved in the ConcePTION ecosystem and show that they are motivated. It was thought that when people show they are actively involved the feeling of doing this together grows, and with that, the willingness to work even harder or more.

One respondent suggested that it could help to work in smaller groups, to stimulate the personal relationships between people and to make it easier for individuals to contribute to the discussion.

Safeguards
Some participants also mentioned safeguards as conditions for working for the ConcePTION ecosystem. A few participants emphasized that the pharmaceutical industry cannot and should not be too involved in the processes of formulating research questions, writing up protocols, and analyzing results. Their involvement could trouble the primary goal of the research, or it was considered too difficult to always align the goal of the private and public industry.

Another safeguard mentioned by a few participants was democratic decision-making regarding the development of scripts and the ability to test scripts and check the results of analyses for inconsistencies. Some participants mentioned that it is important to their organization that they at least feel they can support the results and can take full responsibility for the quality of the data analyses. Lastly, it was mentioned by some participants that they want a clear timeline and communication regarding tasks, deadlines, funding and implementation.

Lastly, one participant also mentioned their obligation towards data providers, and that because of this obligation they would like to remain in control over some of the review processes in terms of data usage and data analyses.

Theme 3: Challenges
During the interviews, many participants discussed current challenges they experience in working for the ConcePTION project. These discussion points can be used for the development of the ConcePTION ecosystem, to overcome these.

Communication and oversight
Although most participants highly value that so many different organizations and institutes are connected to the ConcePTION project, they also mentioned the challenges that come with a project of this magnitude. Many participants mentioned the difficulty in creating clear communication about goals, timelines, and tasks, and emphasized the challenge of not becoming overloaded with emails and other types of digital communication. Having so many different types of digital communication

troubles the overview on tasks etc.

Another challenge many participants mentioned was to keep up with the current state of affairs, progression, task deadlines and above all the work of other people working within the project.

Data (is not information)

A few participants explained that there are also challenges in harmonizing the databases and executing studies because of the heterogeneity of the data across all different databases. Furthermore, organizations have different database and IT-systems that challenge the ability to do studies. Some argued that in order for an ecosystem to sustain, it should have some type of harmonization in the process of data collection, processing, storage and exchange. Before harmonization can be accomplished, it is important for the ConcePTION project to exactly know what data organizations have or have access to, but more importantly: what you can actually do with the data; what type of research questions can be asked for which the data can be of help. With that, some participants also mentioned that it is challenging to generate reliable information based on so many different data, databases, and IT systems. And most importantly: data is not (yet) information. Participants mentioned different practices that could help overcome this challenge of generating reliable information. Often mentioned by the participants as a solution is to have people, who have been working with the data, involved in the translation of the results of the data analyses to make sure data can be translated to the clinicians who need to do something with new findings in clinical practice. These people know the data, the meaning of particular data and the population that is represented in the data. With that, participants argued that if you do not have people involved who know the clinical practice, who are not experienced in the “real life” setting, might lead to having the wrong impression of the data. Second, participants mentioned the need for security/ quality assessments, to make sure scripts fit the data and are correctly run at every organization. Third, participants discussed how working in smaller teams to exchange experiences with tasks, data analyses and research question could help in creating a clearer overview of the possibilities and limitations of databases. Fourth, a less explicit solution that was offered by a few of the participants, is that working with data and aiming towards some level of harmonization demands open-mindedness and trust. Open-mindedness was thought as important, especially regarding the aim and method of the ecosystem. Furthermore, trust was considered to be important for the collaboration between the organizations, but it was also mentioned that there needs to be trust in decisions made by people taking a more leading role in the ecosystem and trust between the public and private sectors is important to make sure actual knowledge is going to be generated within the ecosystem.

Governance procedures

Half of the participants experience challenges due to governance procedures. On the one hand, it was mentioned that these procedures are challenging because countries have different (oftentimes stricter) data privacy rules. On the other hand, it was mentioned that these procedures are challenging because their own company or organization restricts certain (research) activities. Furthermore, some participants explained that there is the problem of extra caution by some organizations. Some participants argued that they experienced that especially in academia, people are extra careful, which creates an extra barrier in collecting, sharing, and analyzing data. Extra caution sometimes leads to having to take extra steps before doing a study, which, according to some participants, hinders the flow of research. Most participants argued it is challenging for the development of an ecosystem, that all organizations have different (governance) procedures: for how to handle data, how to manage a database, how to get authorization/ access to data, and how to manage privacy. Participants agree that having fragmented governance procedures lead to slow processes and unfulfilled opportunities. A clear overview of what can be done with the data could be of great help, according to the participants.

Scientific value

Another challenge mentioned by the participants is creating scientific value and making sure the ecosystem is also there to answer scientific questions. According to some participants, data science is not science per se. Some participants argued that to make it science, there needs to be a scientific

question and there need to be people who help interpret the data. With that, another participant argued that regarding the long-term vision of the ecosystem, there needs to be a distinction between research and development. Research can be explained as finding answers to questions that have not yet been asked, development is finding to already existing questions and developing a way to further do so.

Another challenge mentioned, was the focus of studies within the ecosystem. Currently, some participants experience that there is lot of focus on studies on congenital anomalies. Participants understand that this focus could have to do with the participation of EUROCAT members, however, they feel it is important to also include other fields of interest to make the ConcePTION ecosystem valuable and sustainable.

Aim of the ecosystem

Lastly, it became clear that participants have different views on what the ecosystem should be, and with that, what it's aim should be. First of all, some participants explained that they find it difficult to imagine what the ConcePTION ecosystem is going to be like, because there are not enough results yet showing the potential of the project. A few participants argued that the ConcePTION ecosystem will be something like a data repository which can be used to update labels, request for analyses, and perform feasibility studies with. Another few participants argued the ConcePTION ecosystem is there to learn from real life data and connect real life data to cases. The participants did not further elaborate on how it would exactly work. Most participants agreed, however, that the development of an ecosystem requires time and effort. According to them, it could take another 5-10 years to develop a working ecosystem that is able to, for example, display data in real time, and it could take that long for people to trust the and start sharing more (detailed) data. Many participants seem to think that the current ConcePTION project is highly ambitious, perhaps too ambitious and that the project will not accomplish all of its formulated aims within the 5 years.

Concluding remarks: interpretation and translation

A private public partnership has some major governance challenges. Both concerning the course of the consortium during the 5-year contracted period, as well as the sustainability of the network after that time has ended. From the interviews, a few major challenges and concerns were mentioned that seem important for the success of the consortium and the sustainability of the network. These can be categorized in five themes:

1. **Communication**
    a. Personal relations (direct contact, involvement, smaller groups)
    b. Expectation management (timeline, impact)
    c. Language (goal, database, IT, research, development, science/ scientific)
    d. Trust and transparency

2. **Goal**
    a. Generating reliable information (data is not information, heterogeneity in data, different databases and IT, possibilities)
    b. Align opportunities, possibilities, and interest (what is possible, what are the (shared) opportunities, what are the interest)
        i. Interest: because of opportunity, improving, bigger context, goal, science (not just data science), show capabilities, work together with experts, learn from others, own registries are not enough, requirements (FDA, EMA).
    c. Achievable and realistic: now it is too ambitious

3. **The boundaries of the public private partnership**
    a. There are both limits and opportunities regarding the collaboration between the public and private sector

   b. Creating a vision/ roadmap for this collaboration after the project-time has ended
   c. Having an open mind and creating trust between the public and private sector

4. **Support**
   a. Resources: financial, computing (IT), manpower
   b. Support with: ethics approval, IT, protocol writing, quality checks of data analyses
   c. Permanent staff (EMA or ConcePTION)

5. **Scientific interest**
   a. Quality checks and transparent decision-making regarding research questions and data analysis
   b. Formulating research questions, based on what is possible with the data
   c. Not just data science and not just development
   d. Not just for the industry (tenders)

General Discussion

In this report, we have presented the results of three different studies: a literature study, a case study and an interview study. This report can be seen as a first step towards the development of a governance structure for ConcePTION.

Challenges for the PPP

As became clear from our literature study, much has been written about governance challenges for LHSs, PPPs and large research consortia. On the one hand these challenges concern methodological and ethical aspects of such methods, on the other hand, they concern the effectiveness and benefits of such collaborations. In our view, it is thus extremely important for the ConcePTION consortium to reflect on the effectiveness, the foreseen benefits, and the public interest of the desired ConcePTION ecosystem, before making decisions on how to move forward and especially decide on whether a PPP is the best structure for building an ecosystem of continuous learning.

From the interviews with the DAPs, it became clear that respondents do have different motives and have different views on the added value of the ConcePTION ecosystem. In the current phase of the project, it might be difficult to envision what the ecosystem might look like in the future and how it would work. Before developing a governance framework, it would be good to establish a version of the ConcePTION ecosystem, based on open conversations with relevant stakeholders within the consortium.

As it might be difficult to envision what the ecosystem might look like, respondents also mentioned that it takes time to build these types of systems. According to them, this also means that organizations should accept that it takes time and should invest time and money to encourage this system change. It would also mean that, sometimes, organizations should try to set aside their interests and trust the proposed way forward. Trust seems to be important, but oftentimes lacking. Especially organizations who already have built a system for signaling, analyzing, and reporting serious effects of medication used in pregnancy and breastfeeding, have a more critical stance in the discussion on how to move the ConcePTION project further. These organizations already have built a method, a structure and need to probably change some of their methods in order to fit into the ConcePTION method. With that, there is a lack of trust among public institutions regarding the pharmaceutical industry and their involvement in ConcePTION. Some DAPs are officially constraint by their institution to collaborate with the pharmaceutical industry within consortia and the like or cannot share any data (pseudonymized or not) with the pharmaceutical industry. This is rather difficult for a PPP and as a result might limit the possibility of building an LHS as a PPP.

There is another challenge relating to the purpose limitation of using data DAPs have and/or have access to. Most of the documents mentioned that data can only be requested for research purposes only. The question arises what type of scientific research? Some DAPs mentioned that they would rather not have the pharmaceutical industry involved in the development of research questions, protocols and the execution of data analyses. This would, again, endanger the PPP structure ConcePTION now has. At the same time, the research interests of the public institutions might not always differ. Scientific research could overlap with risk management, which is probably one of the main research focusses of the pharmaceutical industry. For developing a governance structure, it is important to distinguish types of research and to allow organizations to explicitly explain in which projects they like to be or can be involved.

Transparency

Although ConcePTION is not a legal entity, meaning that it has no legal rights and responsibilities and is reliant on contractual implementation from the institutions in which stakeholders are located for key decisions, the governance structure should be clear to all participating organizations in the ecosystem. Good governance requires that all participating organizations know what governance structures and

procedures are in place, what mechanisms for decision-making have been adopted, and where authority and responsibility for different types of actions are located (for example, at the consortium/ management or at the individual organizations) (Morrison, et al., 2020).

This interpretation of transparency relates to the question of what type of support DAPs can receive in terms of governance. As became clear during the interviews, some DAPs would like to gain support in, for example, getting ethics approval. Furthermore, it should be clear for what type of ethico-legal activities DAPs can receive support and for what activities they should consult their own institute. With that, some of the ethico-legal documents can benefit from clarification of certain aspects that are related to governance. For example, relating to data access and confidentiality: the interpretation of anonymization, pseudonymization and which laws and regulation protect what type of data.

Data Access Committee

To overcome certain governance challenges, it could be worthwhile to set up a Data Access Committee (DAC). Depending on what role the DAC has in the ecosystem, it could serve as a safeguard charged with applying rules meant to ensure an ethically permissible balance between data protection and accessibility (GA4GH, 2021). Furthermore, a DAC could help make sure data reuse has potential social value and that there is low risk of foreseeable harms (Cheah & Piasecki, 2020). To promote data sharing and to motive data subjects, DACs should encourage secondary uses that are in line with the interests of data subjects and the organizations that collect data. A DAC or similar type of governance body would review both applications of organization wanting to make use of the ecosystem and should establish frameworks with clear lines of accountability, terms of reference and membership (Cheah & Piasecki, 2020). However, there are no procedural standards that apply across DACs, and therefore, ConcePTION could learn from other initiatives, such as the Global Alliance for Genomics & Health, which has developed a DAC together with guiding principles and a clear policy standards (see table 6)(2021). They have built a framework for responsible sharing of genomic health-related data and have worked on formulating guiding principles and procedural standards for DACs, which are either essential or desired. According to the alliance, their guiding principles and procedural standards should help establish trust in DAC review processes across institutions, repositories, and jurisdictions and thereby promote more efficient, more secure, and more consistent procedures for access to data.

Table 6 Global Alliance for Genomics and Health: Data Access Committee Guiding Principles and Procedural Standards Policy 2021

Guiding principles	**Policy standards**
Make data accessible to advance research and scientific knowledge, as well as to improve health outcomes, through responsible oversight practices Promote health, wellbeing, and the fair distribution of expected benefits, as well as protect against risks Respect the reasonable expectations of data producers and research participants/data subjects, and the communities to which participants/data subjects belong Maintain procedural fairness for applicants seeking access to data	Purpose of the DACs Transparency Terms of reference Standard Operating Procedures (SOPs) Criteria for assessing access applications Progress reporting Data management incidents DACs as living and learning organizations

Soft law and sustainability

It became clear from the ethico-legal documents, that most of the documents DAPs had shared with us were legal document or document that refer to laws and regulations. To build an ethical and sustainable ecosystem of continuous learning, we believe, both DAPs and ConcePTION should strive to also include so-called soft law, or guidelines. Therefore, a governance structure should not only focus on what is obligated by law, but it should also strive to do better ethically. The CIOMS guidelines is a good example of a guideline, that focusses on guiding principles for a governance structure for institutions where data are collected, archived, and analyzed (CIOMS, guideline 12, 2016).

Sustainability was a much-discussed topic during the interviews and is also mentioned as one of the challenges for large research consortia. First needs to be determined what ought to be sustained and in what way. Because some DAPs find it hard to imagine what ConcePTION might look like in the future and because visions on the added value of ConcePTION differs amongst them, it is important to formulate, together with the partnering organizations, what the aim of ConcePTION is and what aspect(s) need to become sustainable.

For ConcePTION to become a sustainable LHS, it is essential, from an ethical point of view, that learning is embedded in the practice of care, while simultaneously, the results of the embedded learning activities improve the practice of care. Some LHSs start as systes where care and research are aligned by more efficient ways in which can be learned from routinely collected data. But when these systems mature as a healthcare system, results that the system creates should also be returned to the target population of the system (Wouters, et al., 2021). For an LHS for pregnant and breastfeeding people, at some point in time, the knowledge created must also be made available for them and their healthcare providers. In that way, results of the embedded learning activities can actually improve the practice of care, from which we can learn subsequently. When LHSs mature, it will be essential that stakeholders recognize and embrace the need for and the added value of the system itself. Beyond a project phase, patients, physicians, scientists, boards of directors, data managers, health institutions, pharmaceutical companies, governments, and others should be convinced of the need of this system as an added or new way forward to create knowledge and to accept the value of the knowledge that this system creates. Pregnant people should be empowered to see the added value of participation in the system and contribute by means of data sharing, physicians should explain why data sharing is important, hospital managers and database owners should help to make the data FAIR and put governance systems in place to maintain trust in the learning health system, health authorities should acknowledge the social value of the knowledge created, especially when an LHS has proven to be a valuable alternative to clinical trials.

Governance principles

From the literature, we already summarized a few important principles that could be relevant for a governance framework for ConcePTION as an LHS. Most of these principles reappeared in both the ethico-legal documents and interviews.

Principle	Comment
Transparency	Transparency in research activities: being open about the process, outcomes, and risks of a study. Also: transparency regarding the type of use and regarding data sharing. Data controllers should provide coherent information to data subjects and should make the effort in making the data Findable, Accessible, Interoperable, and Reusable (FAIR). Transparency in terms of governance: those internal or external to the project know what governance structures and procedures are in place, what mechanisms for legitimate decision-making have been adopted, and where authority and responsibility for different types of actions are located in the consortium or are the responsibility for the individual institutions. A Data Access Committee (DAC) should be installed to provide guidance on conditions for data access, recognition, and supporting agreements.
Privacy and confidentiality	Respect the privacy of data subjects and respect the reasonable expectations of data producers and data subjects and their communities.
Informed consent	Organizations that collect and store health data have a responsibility to make sure informed consent is asked or pregnant and breastfeeding people are actually able to object to the collection of their health data.
Social and scientific value	Promote health and wellbeing and the fair distribution of expected benefits as well as protect against risks. Make data accessible to advance research and scientific knowledge, as well as to improve health outcomes. Maintain procedural fairness for applicants seeking access to data.
Accountability/ Liability	Compliance with obligations provided by the GDPR and national and/or regional legislation. Compliance with guidelines that promote sustainable and ethically responsible health data research. To create a level of accountability for the way respect for autonomy of data subjects is being handled within the LHS, there should be a Data Access Committee (DAC) or other type of governance body installed.
Sustainability and legacy planning	Determine what needs to become sustainable. All stakeholders should accept the system change, and should recognize and embrace the need for and the added value of the LHS. With that, the learning should be embedded in the practice of care, while simultaneously, the results of the embedded learning activities should aim to improve the practice of care. In terms of legacy planning: arrangements should be made regarding: 1) Legal and contractual arrangements after the project; 2) Who has responsibility for co-created data and samples when the project ends and funding ceases; 3) Engagement public and internal (DAPs).
Stakeholder engagement	Include stakeholders in the development of an LHS. Make a

	communication plan, which includes what needs to be communicated to the data subjects and their community, why and where the data subjects need to be engaged and involved, and how this can be achieved. Support researchers who champion engagement and involvement. Achievement of meaningful engagement should result in data subjects feeling empowered.
Trust	Maintain public trust by engaging the data subject in the LHS. Trust is also gained by being trustworthy. Enhance trust between public and private organizations to stimulate the collaboration between the two.
Return of results	Make sure that the learning becomes embedded in the practice of care. Find appropriate mechanisms for keeping data subjects informed of research outcomes.
Ethical oversight	Determine which research activities need ethical oversight and provide support to organizations who have limited experience with receiving ethics approval for research activities.

References

Ballantyne A, Stewart C. Big Data and Public-Private Partnerships in Healthcare and Research: The Application of an Ethics Framework for Big Data in Health and Research. Asian Bioeth Rev. 2019;11(3):315–26.

Botkin JR. Transparency and choice in learning healthcare systems. Learn Heal Syst. 2018;2(1):1–6.

Brody H, Miller FG. The Research-Clinical Practice Distinction, Learning Health Systems, and Relationships. Hastings Cent Rep. 2013;43(5):41–7.

Charmaz K. Constructing Grounded theory: a practical guide through qualitative analysis: SAGE Publications; 2006.

Cheah PY, Piasecki J. Data Access Committees. BMC Med Ethics. 2020;21(1):1–8.

Council for International Organizations of Medical Sciences & World Health Organization. CIOMS Guidelines. International Ethical Guidelines for Health-related Research Involving Humans. [Internet]. 2016. Available from: https://cioms.ch/publications/product/international-ethical-guidelines-for-health-related-research-involving-humans/.

Global Alliance for Genomics and Health. Data Access Committee Guiding Principles and Procedural Standards Policy. [Internet]. 2021. Available from: https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/framework-for-responsible-sharing-of-genomic-and-health-related-data/

Hollestelle MJ, van der Graaf R, Hartman SD, Sturkenboom MCJM, van Delden JJM. A Learning Healthcare System for pregnant and breastfeeding women: what do women during preconception, pregnancy, and nursing think? – A qualitative study. BMC Pregnancy Childbirth [Internet]. 2022;22(1):1–12. Available from: https://doi.org/10.1186/s12884-022-04675-2

Innovative Medicine Initiative ConcePTION. Description: Key objectives of the project. [Internet]. Available from: https://www.imi-conception.eu/ description/.

Jones RD, Krenz C, Griffith KA, Spence R, Bradbury AR, De Vries R, et al. Governance of a Learning Health Care System for Oncology: Patient Recommendations. JCO Oncol Pract. 2020;OP.20.00454.

Kalkman S, Mostert M, Gerlinger C, Van Delden JJM, Van Thiel GJMW. Responsible data sharing in international health research: A systematic review of principles and norms. BMC Med Ethics. 2019;20(1):1–13.

Klijn, E. H. and Teisman, G. R. (2003) Institutional and Strategic Barriers to Public–Private Partnership: An Analysis of Dutch Cases. Public Money and Management, 23:3 pp1–9.

Kuchinke, W, Ohmann, C, Verheij, RA, Veen, EBV & Delaney, BC. Development towards a learning health system—experiences with the privacy protection model of the TRANSFoRm project. *Data Protection on the Move*. Springer, Dordrecht, 2016. 101-134.

Mann SP, Savulescu J, Sahakian BJ. Facilitating the ethical use of health data for the benefit of society: Electronic health records, consent and the duty of easy rescue. Philos Trans R Soc A Math Phys Eng Sci. 2016;374(2083).

Mclennan S, Kahrass H, Wieschowski S, Strech D, Langhof H. The spectrum of ethical issues in a Learning Health Care System: A systematic qualitative review. Int J Qual Heal Care. 2018;30(3):161–8.

Moloney RM, Tambor ES, Tunis SR. Patient and clinician support for the learning healthcare system: Recommendations for enhancing value. J Comp Eff Res. 2016;5(2):123–8.

Morrison M, Mourby M, Gowans H, Coy S, Kaye J. Governance of research consortia: challenges of implementing Responsible Research and Innovation within Europe. Life Sci Soc Policy. 2020;16(1):1–19.

Muller SHA, Kalkman S, van Thiel GJMW, Mostert M, van Delden JJM. The social licence for data-intensive health research: Towards co-creation, public value and trust. BMC Medical Ethics. 2021;22(110): 1-9. https://doi.org/10.1186/s12910-021-006

Olsen, LeighAnne, Dara Aisner, and J. Michael McGinnis. "The learning healthcare system: workshop summary." (2007).

Piasecki J, Dranseika V. Learning to Regulate Learning Healthcare Systems. Cambridge Q Healthc Ethics. 2019;28(2):369–77.

Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. npj Digit Med [Internet]. 2020;3(1):1–7. Available from: http://dx.doi.org/10.1038/s41746-020-00323-1
Rossi A, Lenzini G. Transparency by design in data-informed research: A collection of information design patterns. Comput Law Secur Rev [Internet]. 2020;37:105402. Available from: https://doi.org/10.1016/j.clsr.2020.105402
Seid M, Margolis PA, Opipari-Arrigan L. Engagement, peer production, and the learning healthcare system. JAMA Pediatr. 2014;168(3):201–2.
Spagnuelo D, Ferreira A, Lenzini G. Accomplishing transparency within the general data protection regulation. ICISSP 2019 - Proc 5th Int Conf Inf Syst Secur Priv. 2019;(Icissp):114–25.
Torchia M, Calabrò A, Morner M. Public–Private Partnerships in the Health Care Sector: A systematic review of the literature. Public Manag Rev [Internet]. 2015;17(2):236–61. Available from: http://dx.doi.org/10.1080/14719037.2013.792380
Winter JS, Davidson E. Big data governance of personal health information and challenges to contextual integrity. Inf Soc [Internet]. 2019;35(1):36–51. Available from: https://doi.org/10.1080/01972243.2018.1542648
Wouters RHP, van der Graaf R, Voest EE, Bredenoord AL. Learning health care systems: Highly needed but challenging. Learn Heal Syst. 2020;(June 2019):1–6.
Wouters RHP, van der Graaf R, Rigter T, Bunnik EM, Ploem MC, de Wert GMWR, et al. Towards a responsible transition to learning healthcare systems in precision medicine: Ethical points to consider. J Pers Med. 2021;11(6):1–11.
Xafis V, Schaefer GO, Labude MK, Brassington I, Ballantyne A, Lim HY, et al. An Ethics Framework for Big Data in Health and Research. Asian Bioeth Rev. 2019;11(3):227–54.

Appendix A

Checklist ethico-legal documents

Task: provide ethico-legal documents
What: Checklist task

1. Name organization:				
2. Please indicate the existence and availability of the following documents:				
Document	**Exists** (delete what doesn't apply)	**Available?** (delete what doesn't apply)	**English?**	**Comment**
Policy document(s)	Yes/No	Yes/No	Yes/ No, in:	
Governance structure for data sources	Yes/No	Yes/No	Yes/ No, in:	
Data Processing Agreements (DPAs)	Yes/No	Yes/No	Yes/ No, in:	
Data Access Agreements (DAAs)	Yes/No	Yes/No	Yes/ No, in:	
Data Transfer Agreements (DTAs)	Yes/No	Yes/No	Yes/ No, in:	
Description of approval committees	Yes/No	Yes/No	Yes/ No, in:	
Standard template protocol	Yes/No	Yes/No	Yes/ No, in:	
Study protocol(s)	Yes/No	Yes/No	Yes/ No, in:	
Patient information forms	Yes/No	Yes/No	Yes/ No, in:	
Informed consent forms	Yes/No	Yes/No	Yes/ No, in:	
Other(s), namely: (add more rows for more documents)			Yes/ No, in:	

3. Please indicate whether you have a *Data Privacy Officer*, *Data Controller*, *Data processor* and/or *Data Protection Officer* at your organization and how to contact this person if necessary. **Or please explain how decisions are made regarding data collection, processing and sharing.**

4. Please include a short description below of the requirements of local, organizational and/ or national governance, ethics, scientific or regulatory committees or teams that need to approve study protocols prior to protocol execution in the data sources you will use/access for ConcePTION