

# Data management plan for governance and use of the LifeCycle - EU Child Cohort Network

**Work package 2 - Task 2.2 - Deliverable 2.2**

Based on

Guidelines FAIR Data Management in Horizon 2020 - Version 3.0

Authors:

Evelien van Meel, Liesbeth Duijts, Vincent Jaddoe

Version 1.3

Delivery date: Month 12

Submission date August 2018

## Table of Contents

1. Introduction.....	3
2. Data summary .....	4
3. FAIR Principles .....	9
3.1 Making data findable, including provisions for metadata .....	9
3.2 Making data openly accessible .....	10
3.3 Making data interoperable .....	11
3.4 Increase data re-use (through clarifying licences) .....	12
4. Allocation of resources .....	13
5. Data security.....	13
6. Ethical aspects .....	13
7. General Data Protection Regulation (GDPR).....	13
8. Other.....	14

## 1. Introduction

The LifeCycle Project is designed to establish the EU Child Cohort Network, which brings together existing, successful pregnancy and child cohorts and biobanks, by developing a governance structure taking account of national and European ethical, legal and societal implications, a shared data-management platform and data-harmonization strategies. LifeCycle enriches this EU Child Cohort Network by generating new integrated data on early life stressors related to socio-economic, migration, urban environment and life-style determinants, and capitalize on these data by performing hypothesis-driven research on early life stressors influencing cardio-metabolic, respiratory and mental health trajectories during the full LifeCycle, and the underlying epigenetic mechanisms. LifeCycle translate these results into recommendations for targeted strategies and personalized prediction models to improve health trajectories for current and future Europeans generations by optimizing their earliest phase of life. To strengthen this long-term collaboration, LifeCycle organizes yearly international meetings open to pregnancy and child cohort researchers, introduce a Fellowship Training Programme for exchange of junior researchers between European pregnancy or child cohorts, and develop e-learning modules for researchers performing life course health studies. Ultimately, LifeCycle will lead to a unique sustainable EU Child Cohort Network, and provide recommendations for targeted prevention strategies by identification of novel markers of early life stressors related to health trajectories throughout the LifeCycle.

This data management plan for governance and use of the LifeCycle - EU Child Cohort Network describes the principles of data management, both during and after the project. LifeCycle want to develop a sustainable EU Child Cohort Network, which forms a basis for a long-term collaboration between birth cohorts, beyond the duration of LifeCycle. LifeCycle is not focused on new data-collection within cohorts but on developing an infrastructure for management, access, and safety of existing harmonized data taking account local and international ethical, legal and social principles. This document will be updated at each reporting period.

## 2. Data summary

### a) *State the purpose of the data collection/generation*

LifeCycle is designed to establish the EU Child Cohort Network, which brings together existing, successful pregnancy and child cohorts and biobanks, by developing a governance structure taking account of national and international ethical, legal and societal principles, a shared data management platform and data harmonization strategies. LifeCycle uses only data that is already available within partners and their cohorts. No new data is collected as part of LifeCycle. LifeCycle does not own any data, but is developed to create an infrastructure to bring these data together.

### b) *Explain the relation to the objectives of the project*

The **general objective** of LifeCycle is to bring together pregnancy and child cohort studies into a new, open and sustainable EU Child Cohort Network, to use this network for identification of novel markers of early-life stressors affecting health trajectories throughout the LifeCycle, and translate findings into policy recommendations for stratified and targeted prevention strategies.

**Specific objectives** are to:

- Establish a new, open and long-term multidisciplinary EU Child Cohort Network.
- Enrich the EU Child Cohort Network with novel integrated data on early-life stressors.
- Identify early-life stressors influencing developmental adaptations and risk of NCDs during the LifeCycle.
- Develop analytical methods for causal inference and longitudinal modelling in observational cohort studies.
- Identify DNA methylation markers mediating the relations of early-life stressors and health trajectories.
- Identify interventions on early-life stressors that affect life-course health and develop recommendations.
- Generate personalized risk prediction models in eHealth applications based on early-life stressors.

Research data related to these objectives include data available in pregnancy and birth cohort studies (see 1c)

### c) *Specify the types and formats of data generated/collected*

A list of partners and their cohorts is given in **Table 1**. Research data within these cohorts come from:

- Surveys
- Questionnaires and interviews
- Observations
- Clinical measurements
- Medical records
- Electronic health records

- Administrative records
- Environmental data
- Biological samples
- Genotypic and other omic data
- Other individual data

Data include:

Main exposures: Socio-economic, migration, urban environment, life-style and nutrition related factors, genome wide association screen.

Main mediators: Epigenetics, metabolomics, allergy, brain development

Main outcomes: Cardio-metabolic (BMI, blood pressure, cardiac structure and function, lipids, insulin, glucose); Respiratory (wheezing, infections, lung function, asthma), Mental (behaviour, cognition, education, ASD, ADHD, anxiety, depression).

**Table 2** gives all data per cohort that has been collected previously and will be used for addressing the LifeCycle objectives.

**d) Specify if existing data is being re-used (if any)**

LifeCycle (re-)uses data that is already available within partners and their cohorts. No new data is collected as part of LifeCycle.

**e) Specify the origin of the data**

See 1a and 1 c

**f) State the expected size of the data (if known)**

This is not easy to address since all data belong to specific partners and cohorts.

**g) Outline the data utility: to whom will it be useful**

LifeCycle data will be useful for all junior and senior researchers focused on pregnancy and child health and long term consequences.

**Table 1. List of LifeCycle partners and their cohorts**

No	Participant Organisation Name	Short name	Country	Cohorts
1	Erasmus Universitair Medisch Centrum Rotterdam	ERASMUS	Netherlands	Generation R Study Generation R Next Study
2	Fundació Centre de Recerca en Epidemiologia Ambiental	CREAL	Spain	INfancia y Medio Ambiente Project (INMA)
3	Universita Degli Studi di Torino	UNITO	Italy	Nascita e INFanzia, gli Effetti dell'Ambiente (NINFEA)
4	University of Southampton	UOS	UK	Southampton Women's Survey (SWS)

5	University of Bristol	UNIVBRIS	UK	Avon Longitudinal Study of Parents and Children (ALSPAC / ALSPAC – G2)
6	Kobenhavns Universitet (University of Copenhagen)	UCPH	Denmark	Danish National Birth Cohort (DBC)
7	Bradford Teaching Hospitals NHS Foundation Trust	BTHFT	UK	Born in Bradford (BIB)
8	Academisch Ziekenhuis Groningen (University Medical Center Groningen)	UMCG	Netherlands	Groningen Expertise Centrum voor Kinderen met Overgewicht (GECKO)
9	Panepistimio Kritis (University of Crete)	UOC	Greece	Mother Child Cohort in Crete (RHEA)
10	Nasjonalt Folkehelseinstitutt (Norwegian Institute of Public Health)	NIPH	Norway	Norwegian Mother and Child Cohort (MOBA)
11	Institut National de la Santé et de la Recherche Médicale	INSERM	France	Étude Longitudinale Française depuis l’Enfance) (ELFE / EDEN)
12	Oulun Yliopisto (University of Oulu)	UOULU	Finland	Northern Finland Birth Cohort studies (NFBC66, NFBC86)
13	Samfundet Folkhälsan i Svenska Finland RF	SF	Finland	Helsinki Birth Cohort Study (HBCS)
14	Ludwig-Maximilians-Universität München	LMU	Germany	Childhood Obesity Programme Study (CHOP)
15	Stichting VU-VUMC	VU	Netherlands	Not applicable
16	Slimmere Zorg B.V.	SZORG	Netherlands	Not applicable
17	Concentris Research Management GmbH	CONC	Germany	Not applicable
18	University of Western Australia	UWA	Australia	Western Australian Pregnancy Cohort Study (RAINE)

**Table 2. List of LifeCycle cohorts and their data**

Cohort, N Reference	Design, birth years, Follow-up Main early-life stressors	Available mediators	Available outcomes
<b>ALSPAC</b> N=14,500 <i>Fraser, Int J Epidemiol 2012</i>	Prospective, 1991-1992. Pregnancy – 25 yrs. Socio-economic, migration, and life-style determinants, genome wide association screen.	Epigenetics Metabolomics Allergy Brain development by MRI	<b>Cardio-metabolic:</b> BMI, blood pressure, cardiac structure and function, lipids, insulin, glucose. <b>Respiratory:</b> wheezing, infections, asthma, lung function. <b>Mental:</b> behaviour, cognition, education, ASD, ADHD, anxiety, depression.
<b>ALSPAC-G2</b> N =2,000 <i>Boyd, Int J Epidemiol 2012</i>	Prospective, from 2011. Preconception – 2 yrs. Socio-economic, migration and life-style determinants.	Epigenetics Metabolomics Brain development by ultrasound	<b>Cardio-metabolic:</b> BMI, blood pressure. <b>Respiratory:</b> wheezing, asthma. <b>Mental:</b> behaviour, cognition.

<b>BIB</b> N=11,000 <i>Wright, Int J Epidemiol 2014</i>	Prospective, 2007-2011. Pregnancy – 9 yrs. Socio-economic, migration, urban environment, and life-style determinants, genome wide association screen.	Epigenetics Metabolomics Allergy Brain development by ultrasound	<b>Cardio-metabolic:</b> BMI, blood pressure, lipids, insulin, glucose. <b>Respiratory:</b> wheezing, infections, asthma, lung function. <b>Mental:</b> behaviour, cognition, education, ASD, ADHD, anxiety, depression.
<b>CHOP</b> N=500 <i>Koletzko, Am J Clin Nutr 2009</i>	Prospective, 2002-2004. Pregnancy – 11 yrs. Socio-economic, life-style determinants, genome wide association screen.	Epigenetics Metabolomics, Allergy	<b>Cardio-metabolic:</b> BMI, blood pressure, cardiac structure and function, lipids, insulin, glucose. <b>Respiratory:</b> wheezing, asthma. <b>Mental:</b> behaviour, cognition.
<b>DNBC</b> N=70,000 <i>Olsen, Scand J Public Health 2001</i>	Prospective, 1996-2002. Pre-pregnancy – 20 yrs. Socio-economic, migration, urban environment, and life-style determinants, genome wide association screen.	Allergy	<b>Cardio-metabolic:</b> BMI. <b>Respiratory:</b> wheezing, infections, asthma, lung function. <b>Mental:</b> behaviour, cognition, education, ASD, ADHD, anxiety, depression.
<b>EDEN</b> N=2,000 <i>Drouillet, Br J Nutr 2009</i>	Prospective, 2003-2005. Pre-school – 15 yrs. Socio-economic, migration and life-style determinants.	Allergy	<b>Cardio-metabolic:</b> BMI, blood pressure, lipids, insulin, glucose. <b>Respiratory:</b> wheezing, lung function, asthma. <b>Mental:</b> behaviour, cognition, education.
<b>ELFE</b> N=18,000 <i>Vandentorren, BMC Pediatr 2009</i>	Prospective, 2011. Pre-school – 5 yrs. Socio-economic, migration, urban environment, life-style determinants	Allergy	<b>Cardio-metabolic:</b> BMI. <b>Respiratory:</b> wheezing, infections, asthma. <b>Mental:</b> behaviour, cognition.
<b>GECKO</b> N=2,500 <i>L'Abée, Int J Epidemiol 2009</i>	Prospective, 2006-2007. Pregnancy – 10 yrs. Socio-economic, migration, life-style	Allergy	<b>Cardio-metabolic:</b> BMI, blood pressure. <b>Respiratory:</b> wheezing, asthma. <b>Mental:</b> behaviour, education.
<b>Generation R</b> N=7,000 <i>Jaddoe, Eur J Epidemiol 2012</i>	Prospective, 2002 – 2006. Pregnancy – 15 yrs Socio-economic, migration, urban environment, and life-style determinants, genome wide association screen.	Epigenetics Metabolomics Allergy Brain development by ultrasound/MRI	<b>Cardio-metabolic:</b> BMI, blood pressure, cardiac structure and function, lipids, insulin, glucose. <b>Respiratory:</b> wheezing, infections, lung function, asthma. <b>Mental:</b> behaviour, cognition, education, ASD, ADHD, anxiety, depression.
<b>Generation R Next</b> N=2,000 <i>Jaddoe, Eur J Epidemiol 2012</i>	Prospective, 2016-2019. Pre-pregnancy – 2 yrs. Socio-economic, migration, urban environment, and life-style determinants.	Epigenetics Metabolomics Allergy Brain development by ultrasound/MRI	<b>Cardio-metabolic:</b> body mass index, blood pressure, cardiac structure and function, lipids, insulin, glucose. <b>Respiratory:</b> wheezing, infections, lung function, asthma. <b>Mental:</b> behaviour, cognition,
<b>HBCS</b> N=13,000 <i>Eriksson, BMJ 2001</i>	Prospective, 1934 – 1944. Pregnancy – 80 yrs. Socio-economic, migration, and life-style determinants, genome wide association screen.		<b>Cardio-metabolic:</b> BMI, blood pressure, lipids, insulin, glucose, hypertension, type 2 diabetes, dyslipidaemia. <b>Respiratory:</b> asthma, COPD. <b>Mental:</b> cognition, psychiatric illness.

<b>INMA</b> N=3,500 <i>Guxens, Int J Epidemiol 2011</i>	Prospective, 1997-2008. Pregnancy – 20 yrs. Socio-economic, migration, urban environment, and life-style determinants, genome wide association screen.	Epigenetics Allergy Brain development by ultrasound/MRI	<b>Cardio-metabolic:</b> BMI, blood pressure, lipids, insulin, glucose. <b>Respiratory:</b> wheezing, respiratory infections, lung function, asthma. <b>Mental:</b> behaviour, cognition, ASD, ADHD, anxiety, depression.
<b>MoBa</b> N=90,000 <i>Magnus, Int J Epidemiol 2006</i>	Prospective, 1999-2008. Pregnancy – 17 yrs. Socio-economic, migration, and life-style determinants, genome wide association screen.	Epigenetics Allergy	<b>Cardio-metabolic:</b> BMI, blood pressure. <b>Respiratory:</b> wheezing, respiratory infections, lung function, asthma. <b>Mental:</b> behaviour, cognition, ASD, ADHD, anxiety, depression.
<b>NFBC1966</b> N=12,000 <i>Jarvelin, Hypertension 2004</i>	Prospective, 1966. Pregnancy – 50 yrs. Socio-economic, migration, life-style determinants, genome wide association screen.	Epigenetics Metabolomics Allergy Brain development by MRI	<b>Cardio-metabolic:</b> BMI, blood pressure, cardiac structure and function, lipids, insulin, glucose. <b>Respiratory:</b> wheezing, respiratory infections, lung function, asthma, COPD. <b>Mental:</b> behaviour, cognition, education, ASD, ADHD, anxiety, depression.
<b>NFBC1986</b> N=9,500 <i>Jarvelin, BJOG 1993</i>	Prospective, 1986. Pregnancy – 30 yrs. Socio-economic, migration, urban environment, and life-style determinants, genome wide association screen.	Epigenetics Metabolomics Allergy Brain development by MRI	<b>Cardio-metabolic:</b> BMI, blood pressure, cardiac structure and function, lipids, insulin, glucose. <b>Respiratory:</b> wheezing, respiratory infections, lung function, asthma, COPD. <b>Mental:</b> behaviour, cognition, education, ASD, ADHD, anxiety, depression.
<b>NINFEA</b> N=7,500 <i>Richiardi, Eur J Epidemiol 2007</i>	Prospective, 2005 – 2016. Pregnancy – 11 yrs. Socio-economic, urban environment, and life-style determinants.	Allergy	<b>Cardio-metabolic:</b> BMI. <b>Respiratory:</b> wheezing, respiratory infections, asthma. <b>Mental:</b> behaviour, education.
<b>RAINE</b> N=2,900 <i>Newnham, Lancet 1993</i>	Prospective, 1989-1992. Pregnancy – 25 yrs. Socio-economic, migration, urban environment, and life-style determinants, genome wide association screen.	Epigenetics Metabolomics Allergy Brain development	<b>Cardio-metabolic:</b> BMI, blood pressure, lipids, insulin, glucose. <b>Respiratory:</b> wheezing, respiratory infections, lung function, asthma. <b>Mental:</b> behaviour, cognition, education, ASD, ADHD, anxiety, depression.
<b>RHEA</b> N=1,300 <i>Chatzi, Am J Epidemiol 2009</i>	Prospective, 2007 – 2008. Pregnancy – 7 yrs. Socio-economic, migration, urban environment, and life-style determinants.	Epigenetics Metabolomics Allergy	<b>Cardio-metabolic:</b> BMI, blood pressure, lipids, insulin, glucose. <b>Respiratory:</b> wheezing, respiratory infections, lung function, asthma. <b>Mental:</b> behaviour, cognition, education, ASD, ADHD, anxiety, depression.
<b>SWS</b> N=3,000 <i>Inskip, Int J Epidemiol 2006</i>	Prospective, 1998-2007. Prepregnancy – 18 yrs. Socio-economic, migration, and life-style determinants.	Allergy	<b>Cardio-metabolic:</b> BMI, blood pressure, lipids, insulin, glucose. <b>Mental:</b> behaviour, cognition, education, anxiety, depression.



### 3. FAIR Principles

#### 3.1 Making data findable, including provisions for metadata

In general data related to LifeCycle is findable through different websites. Links to these websites will be available on the LifeCycle website (and later EU Child Cohort Network website)

- ***www.birthcohorts.net***
  - Overview of all cohorts and their data
  - Open website, no restriction for access on cohort information
  - Find function is included in website
- ***EU Child Cohort Network Variable Catalogue***
  - Overview of harmonized data and variables in each cohort
  - In first phase access for LifeCycle partners, when finished available for cohort researchers
  - Find function is included in website
- ***LifeCycle Data***
  - Harmonized data from different cohorts
  - Data server is within institutional firewall
  - Access to data can only be given by data owner (LifeCycle partner)
  - Enables data analyses

#### ***a) Outline the discoverability of data (metadata provision)***

LifeCycle has revitalized the [www.birthcohorts.net](http://www.birthcohorts.net) website. This website gives an overview of available pregnancy and birth cohorts and the data available in these cohorts. More specific details of data available in cohorts and their main characteristics will be available in a data dictionary. As part of the LifeCycle project, the EU Child Cohort Network Variable Catalogue has been developed. The aim of the catalogue is to provide an overview of available data in the EU Child Cohort Network in order to facilitate long-term collaborations between existing pregnancy and child cohorts. The catalogue was built using the MOLGENIS software platform for scientific data (<http://molgenis.org>, Swertz et al, Bioinformatics, 2014). The catalogue is an openly and accessible online platform that documents what type of harmonized variables (LifeCycle variables) are available from each participating cohort. It also documents how each cohort has harmonized these variables, including the raw variables (source variables) used by the cohorts, which were used to derive the LifeCycle variables. No actual data are contained in the online catalogue. The catalogue also contains a description of each cohort included in the LifeCycle project with links to the code books.

#### ***b) Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?***

All data related to individuals will be available only coded with participant unique identifiers, which are not able to link data to individuals. These identifiers are cohort specific. Since LifeCycle works with a federal analysis tool (DataSHIELD), all data will remain in the specific centers.

#### ***c) Outline naming conventions used***

All harmonized data from different partners and cohorts will be renamed, with using the same names between these different cohorts. Since the data structure is still being developed no clear examples are available yet.

**d) Outline the approach towards search keyword**

The EU Child Cohort Network Variable Catalogue will contain a search option to find relevant data

**e) Outline the approach for clear versioning**

The LifeCycle and EU Child Cohort website will be used to present the cohorts ([www.birthcohorts.net](http://www.birthcohorts.net)) and the new developed EU Child Cohort Network Variable Catalogue.

**f) Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how.**

See 2.1.a-e

### **3.2 Making data openly accessible**

**a) Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**

LifeCycle researchers will use existing data from different partners and their cohorts. All data have already been collected and are available for scientific research. Now new data are collected as part of LifeCycle. All research data relevant for the LifeCycle objectives are available. Data to address other non-LifeCycle related research questions can be requested at each specific LifeCycle partner. Availability of data is related to local and international ethical, social and legal principles. LifeCycle builds an infrastructure which enable safe data sharing with other partners.

**b) Specify how the data will be made available**

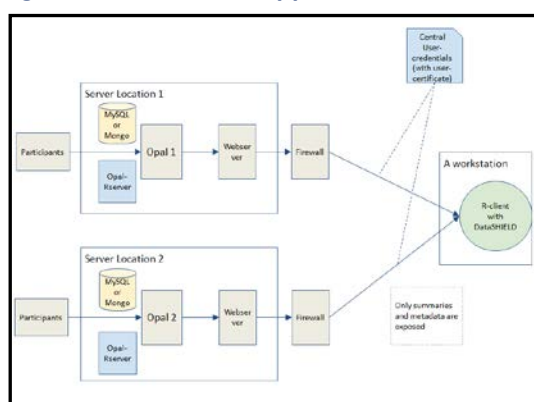
The data management and access is based on the following key principles:

- Full compliance with best practice in data privacy and security.
- Use of coded data with consent.
- Use of privacy enhancing technologies such as filters.
- Use of policies that will be implemented to enable greater use of data in research.
- Best practice in the security of infrastructure and data assets.
- All procedures, policies and methods are approved by the relevant local authorities.

Management and access of all data is the primary responsibility of each partner and their cohort, and is performed according to relevant local and international principles. Data relevant for LifeCycle will be made available by an additional server located at each specific partner. These additional Opal servers enable access from other institution, conditional on approval of data access. The general data-management structure will be based on the open DataSHIELD platform, which enables pooling of harmonized data without physical transfer of data between partners (**Figure 1**). The open DataSHIELD platform, which has been developed as part of the EU FP7-funded BioSHaRE project, is used as a basis for the data-management platform. Specific data

servers and IT support will be provided by LifeCycle. Within the DataSHIELD platform, each partner (and cohort) remains responsible for its own data safety, harmonisation, sharing or access according to any approved local ethical, legal and social principle. Data access to a partner specific server is given by the partner and is subject to a mutual data sharing agreement. In principle, LifeCycle will perform research based on individual participant data from all these different cohorts using the DataSHIELD platform. However, partners and their cohorts can always decide to share research data without using DataSHIELD, conditional on relevant local ethical, legal and social approvals a mutual data sharing agreement.

**Figure 1: DataSHIELD approach**



c) **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**

See 2.2.b

d) **Specify where the data and associated metadata, documentation and code are deposited**

See 2.1. and 2.2.b

e) **Specify how access will be provided in case there are any restrictions**

See 2.2.a

### **3.3 Making data interoperable**

a) **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**

LifeCycle researchers will use existing data from different partners and their cohorts. To perform analyses across cohorts, existing data is being harmonized and integrated exposure variables are being developed within each cohort. Strategies for harmonization (approaches, syntax, outcomes definitions) are developed within the consortium and will be available upon request when the harmonisation is finished. For details about availability of data and metadata 2.2.

- b) Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

Details of data available in cohorts and their main characteristics will be available in a website data dictionary (EU Child Cohort Network Variable Catalogue). Links to these websites will be available on the LifeCycle website. The vocabulary for all data types will be standardized between cohorts to allow inter-disciplinary interoperability.

### **3.4 Increase data re-use (through clarifying licenses)**

- a) Specify how the data will be licensed to permit the widest reuse possible**

LifeCycle uses only data that is already available within partners and their cohorts. No new data is collected as part of LifeCycle. LifeCycle does not own any data, but is developed to create an infrastructure to bring these data together. LifeCycle will develop a safe, robust and up-to-date data sharing platform (DataSHIELD), which links harmonized data from different partners and their cohorts within and outside LifeCycle.

- b) Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**

The data will be available for (re)-use when the DataSHIELD network is available and the data have been harmonized. This will be completely available at the end of LifeCycle.

- c) Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**
- LifeCycle will develop a safe, robust and up-to-date data sharing platform (DataSHIELD), which links harmonized data from different partners and their cohorts. The platform and harmonized data will be available after the end of the project. Permission should be given by each partner who owns the data.

- d) Describe data quality assurance processes**

Each partner is responsible for the quality of the primary, already available, data collection in the own cohort. The LifeCycle data harmonization is performed by a data-harmonization group with representatives of each partner. This group has regular meetings and report back to the full consortium at each LifeCycle General Assembly. Harmonized data are made available and shared via the DataSHIELD platform.

- e) Specify the length of time for which the data will remain re-usable**

Data will at least be available for the duration of LifeCycle. However, LifeCycle partners are keen to keep the structure of sharing harmonized data via the DataSHIELD platform available for other projects. LifeCycle uses the DataSHIELD platform because of the free availability. Also, this approach also enable partners to use their data and server for other future collaborations.

#### 4. Allocation of resources

**a) Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**

Costs for servers for servers and data-management are within the budget of each partner.

**b) Clearly identify responsibilities for data management in your project**

Each partner is responsible for the data storage, quality, management and sharing according to local and international ethical, legal and social principles and according to the LifeCycle objectives. LifeCycle partners are together responsible for developing and integrating harmonized variables and setting up a specific server for data-sharing. It is the role of LifeCycle Executive Board (all WP leads) to define the consortium's policy in respect of Information Governance, taking into account legal and individual institutional requirements. LifeCycle coordinator and co-investigators are also responsible for ensuring that sufficient resources are provided to support the policy requirements.

**c) Describe costs and potential value of long term preservation**

After LifeCycle, the data sharing platform and harmonized data will remain available. However, updating of data or servers will probably require additional budgets.

#### 5. Data security

**a) Address data recovery as well as secure storage and transfer of sensitive data**

Each partner is responsible for the data storage, quality, management and sharing according to local and international ethical, legal and social principles and according to the LifeCycle objectives. Policies for data security are locally determined and in line with specific institutional practices. LifeCycle does share data, but does not own data.

#### 6. Ethical aspects

**a) To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former.**

All cohorts have their own responsibility for local ethical and participant consent. Only coded data are shared. All needed ethical approvals from the institution and participant for relevant data collection, use and sharing are the responsibility of the partners. The LifeCycle coordination team has copies of these approvals. All ethical approvals of partners and their cohorts are available upon request. The ownership and primary responsibility of data are always with the institution, which runs its own cohort data collection.

#### 7. General Data Protection Regulation (GDPR).

After four years of preparation and debate the GDPR was finally approved by the EU Parliament on 14 April 2016 with an enforcement date 25 May 2018. The EU General Data Protection Regulation (GDPR) replaces the Data Protection Directive 95/46/EC and was designed to

harmonize data privacy laws across Europe, to protect and empower all EU citizens data privacy and to reshape the way organizations across the region approach data privacy. The key articles of the GDPR, as well as information on its business impact, can be found throughout this site.

LifeCycle was written before the GDPR was active. LifeCycle partners were not familiar with the impact of the GDPR on the planned work. The potential impact of the GDPR on LifeCycle activities has been and will be discussed as part of the WP2 discussions focused on the data management and governance structure.

LifeCycle aims to set up a federal data analysis infrastructure, in which the data remains within the institute that owns the data. Management and access of all data is the primary responsibility of each partner and their cohort, and is performed according to relevant local and international principles, including the GDPR. Data relevant for LifeCycle will be made available by an additional server located at each specific partner. Within the DataSHIELD platform, each partner (and cohort) remains responsible for its own data safety, harmonisation, sharing or access according to any approved local ethical, legal and social principles. Data access to a partner specific server is given by the partner and is subject to a mutual data sharing agreement. These data sharing agreements has to be in line with the GDPR. In principle, LifeCycle will perform research based on individual participant data from all these different cohorts using the DataSHIELD platform.

Because of the federal data analysis platform, the direct implications of the GDPR on LifeCycle research projects seems to be small. However, the potential consequences will be discussed at each coming General Assembly meeting. We will invite experts from partners to discuss whether our approach is in line with the GDPR.

## 8. Other

**a) *Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)***

Not applicable