



# Lifebrain

---

## **D2.1. Data harmonisation guidelines and standardized data management/processing protocol across sites**

Project title:	Healthy minds from 0-100 years: Optimising the use of European brain imaging cohorts
Due date of deliverable:	30 <sup>th</sup> June, 2017
Submission date of deliverable:	30 <sup>th</sup> June, 2017
Leader for this deliverable:	William Baaré (REGIONH)



This project has received funding from the *European Union's Horizon2020 research and innovation programme* under grant agreement No 732592.

Contributors to the deliverable	Name	Organisation	Role / Title
Deliverable leader	William Baaré	REGIONH	WP2 Leader. Coordination. HUBU study data organization/ codebooks/ shared variables/ variables of initial interest to lifebrain. Quality control, organization, Lifebrain site codebooks. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium.
Co-authors	Kathrine Skak Madsen	REGIONH	HUBU study data organization/ codebooks/ shared variables/ variables of initial interest to lifebrain. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium.
	Louise Baruël Johansen	REGIONH	Quality control, organization, Lifebrain site codebooks. Discussions on Lifebrain data harmonization.
	Olga Rigina	REGIONH	Quality control, organization, Lifebrain site codebooks. Discussions on Lifebrain data harmonization.
	Hartwig Siebner	REGIONH/ University of Copenhagen	Overview and feedback on harmonization of LISA study data.
	Michael Kjær	REGIONH/ University of Copenhagen	Overview and feedback on harmonization of LISA study data.
	Erik Lykke Mortensen	University of Copenhagen, Public health	Overview and feedback on harmonization of LISA study data.
	Ellen Garde	University of Copenhagen, Public health/ REGIONH	Overview and feedback on harmonization of LISA study data. LISA study data organization/ codebooks/ shared variables
	Christian Eriksen	REGIONH	LISA study data organization/ codebooks/ shared variables
	Kristine Beate Walhovd	UiO	Overview and feedback on harmonization of LCBC study data. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium
	Anders Martin Fjell	UiO	Overview and feedback on harmonization of LCBC study data. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium
	Monika Mowinckel	UiO	LCBC study data organization/ codebooks/ shared variables
	Rene Westerhausen	UiO	Overview and feedback on harmonization of LCBC study data. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium

	Lars Nyberg	Umeå	Overview and feedback on harmonization of Betula study data. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium
	Mikael Stiernstedt	Umeå	Overview and feedback on harmonization of Betula study data. Betula study data organization/ codebooks/ shared variables/ variables of initial interest to Lifebrain.
	Klaus Ebmeier	UOXF	Overview and feedback on harmonization of White hall studies data. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium
	Eniko Zsoldos	UOXF	Oxford Imaging data organization/ codebooks/ shared variables/ variables of initial interest to lifebrain. Guidance on Whitehall II data overseen by Aida Sanchez
	Aida Snachez	UOXF/UCL	Whitehall II Senior Data Manager. Whitehall II data organization/ codebooks/ shared variables/ variables of initial interest to lifebrain
	Ulman Lindenberger	MPIB	Overview and feedback on harmonization of BASE and BASE-II data. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium.
	Andreas Brandmaier	MPIB	Statistical advisory contributions to BASE and BASE-II studies. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium.
	Simone Kühn	MPIB	BASE-II imaging data organization/ codebooks/ shared variables/ variables of initial interest to Lifebrain. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium.
	Sandra Düzel	MPIB	BASE-II data organization/ codebooks/ shared variables/ variables of initial interest to Lifebrain. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium.
	Julia Delius	MPIB	BASE data organization/ codebooks/ shared variables/ variables of initial interest to Lifebrain. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium.
	Martin Becker	MPIB	Compilation of BASE variables for codebooks.
	David Bartrés Faz	UB	Overview and feedback on harmonization of the University of Barcelona studies data. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium
	Cristina Solé Padullés	UB	University of Barcelona studies data organization/ codebooks/ shared variables/ variables of initial interest to Lifebrain.
	Brenda Penninx	VMUC	Overview and feedback on harmonization of NESDA study data. NESDA study data organization/ codebooks/ shared variables/ variables of initial interest to lifebrain. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium
	Laura Nawijn	VMUC	NESDA data organization/ codebooks/ shared variables/ variables of initial interest to lifebrain. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium
	Rik Henson	MRC/UCAM	Overview and feedback on harmonization of CamCAN study data. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium

	Rogier Kievit	MRC/UCAM	Overview and feedback on harmonization of CALM study data. CALM study data organization/ codebooks/ shared variables/ variables of initial interest to lifebrain. Discussions on which variables to harmonize and how, locally and in the Lifebrain consortium
	Tina Emery	MRC/UCAM	CamCAN and CALM data organization/ codebooks/ shared variables / variables of initial interest to Lifebrain.
	Darren Price	MRC/UCAM	Assistance with CamCAN variables code book
	Annie Bryant	MRC/UCAM	Assistance with CALM variables code book
	Paolo Ghisletta	UNIGE	Advisor for data analyses. Discussions on Lifebrain data harmonization.
Final approval	Barbara B. Friedman	UiO	Final editing, approval and submission

Document History				
Release	Date	Reason for Change	Status (Draft/In-review/Submitted)	Distribution
1.0.	05.06.2017	First draft	Sent to review to the CO	Email
2.0.	28.06.2017	Second draft	Revision	Email
3.0.	30.06.2017	Final version	Submission to the PP	Participant portal

Dissemination level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

## Table of contents

<b>TABLE OF CONTENTS .....</b>	<b>4</b>
<b>EXECUTIVE SUMMARY .....</b>	<b>5</b>
<b>LIST OF ACRONYMS / ABBREVIATIONS.....</b>	<b>7</b>
<b>INTRODUCTION .....</b>	<b>8</b>
TASK 2.1 DESCRIPTION.....	8
BACKGROUND .....	8
STATUS OF TASK/STATE-OF-THE ART IN THE TOPIC.....	9
OBJECTIVES.....	9
PERSPECTIVE .....	10
COLLABORATION AMONG PARTNERS.....	10
<b>1. INCLUDED STUDIES PER SITE .....</b>	<b>11</b>
<b>2. DESCRIPTION OF ACTIVITIES.....</b>	<b>12</b>
1.1. MAPPING THE DATA .....	12
1.2. SELECTION OF VARIABLES OF INITIAL FOCUS .....	14
1.3. GATHERING STUDY PROTOCOLS AND SITE-SPECIFIC VARIABLES META DATA .....	15
1.4. CHECKING CODE BOOK CONSISTENCY .....	16
1.5. PRELIMINARY DATA HARMONIZATION POTENTIAL .....	17
<b>2. NEXT STEPS.....</b>	<b>20</b>
2.1. GENERAL HARMONIZATION PROCEDURES .....	21
2.2. DESIGNING THE LIFE BRAIN DATABASE.....	22
<b>3. CONCLUSION.....</b>	<b>22</b>
<b>4. REFERENCES.....</b>	<b>22</b>
<b>5. APPENDIX .....</b>	<b>26</b>

## Executive Summary

The deliverable is a report on the implementation of part of Task WP2.1 “Data harmonization guidelines and standardized data management plan/processing protocol across sites”.

This task aims to develop and implement harmonization protocols to unify and bridge the different coding standards of target variables between sites. In this process, we will develop guidelines for addressing differences in data management systems and routines, and MR image quality control procedures across sites to secure efficient application and utilization of data transfer, storage and processing systems.

In this deliverable we

1. Categorized all available and in principle shareable data for all Lifebrain studies and time points.
2. Provided necessary information to the work done in other WP2 tasks:
  - a. Task 2.3 “Enrichment of existing cohorts by online data collection (M6-M18)
  - b. Task 2.4 “Enrichment of existing cohorts by home collection of dried blood spots (DBS)” (M6–M36)
  - c. Task 2.5 “Enrichment of existing cohorts by administration of buccal swabs for epigenetic analysis” (M6–M36)
  - d. Task 2.6 “Generation of new GWAS (genome-wide association study) data and harmonization of genetic data across samples” (M6-M24)
  - e. Task 3.2. “Development and validation of a posteriori harmonization of MRI data”, (M1-M9)
3. Identified and selected variables of initial interest
4. Obtained site-specific code books, containing information on inclusion/exclusion criteria, standard operating procedures, data collection methods, and metadata on the variables of initial interest.
5. Provided a general evaluation of the data harmonization potential of variables of initial interest and formulate general data harmonization strategies

Next steps:

1. Continue with quality control and completion of site-specific study protocols.  
The latter specifically concerns documenting the instructions given to subjects during cognitive tests.
2. Continue with the stringent quality control of the site-specific variable code books on completeness, organization and formatting.
3. Establish work groups for a in depth evaluation of data harmonization for specific dimensions and clusters of variables

4. Define and establish variables codes and meta-data for variables of initial interest to Lifebrain. Generate Lifebrain formatted tables template for site-specific variables to the Lifebrain database, developed in Task 3.1 “Development of a data storage and management system” (M1-M6), and develop conversion scripts to populate Lifebrain tables with site-specific data.

## List of acronyms / abbreviations

Lifebrain	Healthy minds from 0-100 years: Optimising the use of European brain imaging cohorts
WP	Work package
GA	Grant Agreement
UiO	University of Oslo
UmU	Umeå Universitet
UOXF	University of Oxford
MPiB	Max Plank Institute for Human Development
UB	University of Barcelona
REGIONH	Region Hovedstaden
VUMC	University Medical Center Amsterdam
MRC	Medical Research Council
UCAM	University of Cambridge
MRI	Magnetic Resonance Imaging

Abbreviations of test and questionnaires that have been employed in Lifebrain studies are provided in Appendix 1.



## Introduction

### Task 2.1 description

*“Task 2.1 Establish a data management and processing protocol for harmonization of data for pooling, integration, quality control across sites. Lead: REGIONH; Participants: UmU, UOXF, MPIB, UB, MRC, VUMC (M1-M6) The task will develop and implement harmonization protocols, i.e. develop technical and semantic data standards, and quality protocols to unify and bridge the different coding standards of target variables between sites. It will also define guidelines*

*for addressing differences in data management systems and routines, and MR image quality control procedures across sites to secure the efficient application and utilization of the data transfer, storage and processing system developed in T3.1. Moreover, T2.1 will facilitate the implementation of the innovative post hoc harmonization algorithms, developed in T3.2, to correct MR images for scanner-specific differences in tissue contrast and intensity, non-uniformity. Finally, T 2.1 will ensure the efficient implementation and management of the new dried blood spot assays and online data collection system developed in T3.5 and 3.6, respectively.”*

### Background

Lifebrain is a European multi-site study aimed to capitalize on existing multidimensional cross sectional and longitudinal retrospectively acquired biological, behavioural and cognitive, medical, mental health, and in vivo neuroimaging data.

The objective of Lifebrain is to establish a solid foundation of knowledge for understanding how brain, cognitive and mental health can be *optimized through the lifespan*. Identification of determinants of brain, cognitive and mental health at different stages of life requires a large database of detailed information about brain imaging measures, cognitive function, mental and medical health, as well as biological measures.

Central to (large-scale) multi-site initiatives such as Lifebrain are data integration, standardization and harmonization. The international European aspect adds an extra layer of unique possibilities as well as challenges. The coordinated research efforts within Lifebrain allows for multiple analytic strategies (2), increase statistical power, allow more advanced subgroup analyses, enhance generalizability of findings and support cross validation or replication of findings across datasets.

## Status of task/state-of-the art in the topic

In June 2016, the international Journal of Epidemiology published the “Maelstrom Research guidelines for rigorous retrospective data harmonization” (2). Over a nine-year period, using *“a phone survey with 34 major international research initiatives, a series of workshops with experts, and case studies applying the proposed guidelines”* the authors developed generic guidelines for successful retrospective data harmonization that include several interdependent steps: “

1. Define the research question, objectives and protocol
- 2. Assemble pre-existing knowledge and select studies**
- 3. Define targeted variables and evaluate harmonization potential**
4. Process data
5. Estimate quality of the harmonized dataset(s) generated
6. Disseminate and preserve final harmonization products”

The current deliverable primarily concerns steps 2 and 3. Overall Lifebrain research questions have been outlined in the grant application and are detailed in WP4: Demonstration. Steps 4, 5 and 6 are out of the scope of the current deliverable and are the main subject of other work packages.

## Objectives

The overall objective of the deliverable is to provide a solid foundation for establishing a Lifebrain database consisting of available behavioural, cognitive, medical and mental health, neuroimaging and biological data, in-depth data harmonization and development and implementation of guidelines for standardized data management and processing protocols across sites.

We will:

- Categorize available data for all Lifebrain studies and time points
- Identify an initial set of variables of interest
- Acquire site-specific code books containing information on study design, standard operating procedures, data collection methods, and metadata on variables e.g. variable names, variable description, data type, data format, allowable values etc.
- Identify commonalities in available data across sites, evaluate general harmonization potential and formulate general data harmonization strategies

## Perspective

Details and implementation of exact data harmonization techniques and the evaluation of their applicability and quality are dependent on the specific research questions formulated in WP4 and what studies will be targeted

Data harmonization, data management and proper documentation will require constant focus and effort throughout the lifespan of Lifebrain.

## Collaboration among partners

The large number of Lifebrain researchers that have contributed to the deliverable reflects the strong, transparent and focussed collaborative framework that Lifebrain provides. Each site designated a team consisting of the study PIs and dedicated research personnel to discuss and specify the dimensions and variables of interest as well as data harmonization and to provide detailed information on study setup and protocols, and metadata for variables of interest as well as expertise and experience in data organization and harmonization.

## 1. Included studies per site

UiO	LCBC: Lifespan Changes in Brain and Cognition (consist of subject of Hukommelse Prosjekt, Neurocognitive Development, Neurocognitive Plasticity, and Mor og Barn) (3 time points)
UmU	Betula - Aging, Memory and Dementia (2 time points)
UOXF	Whitehall II: Imaging (1 time point)
MPIB	BASE (Berlin Altersstudie, 8 time points) BASE II (Berlin Altersstudie, 2 time points)
UB	WAHA (Walnut intervention study: 3 time points) MSA (Multi-Systemic Atrophy: 1 time point) PD (Parkinson's Disease: 3 time points) GABA (GABA Spectroscopy: 1 time point) iTBS (Theta-burst stimulation: 1 time point) CR (Cognitive Reserve: 1 time point)
REGIONH	HUBU ("Hjernens Udvikling hos Børn og Unge"; Brain maturation in children and adolescents: 12 time points) LISA (Live Active Successful Ageing: 1 time point)
VUMC	NESDA (Netherlands Study of Depression and Anxiety: 6 time points)
MRC	CamCan (Cambridge Centre for Ageing and Neuroscience: 3 time points)
UCAM	CALM (Centre for Attention, Learning and Memory: 1 time point)

In Figure 1. we provide the total number of subjects and observations. A complete overview of the total number of unique participants and observations with behavioural data only and with behavioural and MRI data for all individual Lifebrain studies is provided in Appendix 2.

	Lifebrain
	All studies
Total unique subjects	12527
Behaviour only	7327
MRI+Behavior	5200
Total observations	27972
Behaviour only	19873
MRI+Behavior	8099

FIGURE 1. NUMBER OF SUBJECTS AND OBSERVATIONS IN LIFE BRAIN

## 2. Description of activities

The groundwork for the Lifebrain data harmonization effort was laid down in the grant application and during the Lifebrain Kick-off meeting held in Brussels, January 16-18, 2017. During the first 6 months of Lifebrain, the WP2.1 leader, RegionH, and data harmonization teams of individual Lifebrain sites communicated through email, Slack, OneDrive, Skype and telephone to discuss and organize all necessary information.

### 1.1. Mapping the data

The first task was to map all data available and in principle shareable within Lifebrain. Based on the discussion at the kick off meeting a spread sheet was created containing identified dimensions and variables. This spread sheet was then populated and if necessary extended by each site with information on data availability, and used tests, protocols, questionnaires etc. The dimensions that were included are displayed in Figure 2. A detailed overview of the variables within each of dimensions is provided in Appendix 3.

## Healthy minds from 0 to 100 years: Optimising the use of European brain imaging cohorts

Study set up/ participants	Physical measures	Medical Health	Biomarkers
<b>Setup</b>	Weight	physical Examination	Blood
Start and end year	Weight at birth	info on general health	Freezer: -20 / -80
Time between subsequent assessments	Height	info on cardio vascular disease	Whole blood
Assessment date	Grip strength	info on Vascular risk factors	Serum
<b>Participants</b>	Waist circumference	info on stroke	Plasma
Number of subjects	Blood pressure	info on diabetes	Buffy coat
<b>Sex</b>	Handedness	info on Longstanding illness	Info on time of the day
Number of Females	Dexterity	info on Sickness absence	Info on fasting or no fasting
Number of Males	Hearing	Women's health	Info hours since last meal
<b>Age related</b>	Vision	<b>Cognitive Tests</b>	Info on time to Scan/Assessment
Age Range	Eye dominance	IQ or equivalent	Saliva
Can share birth date	Chair rise	Attention	Freezer: -20 / -80
Can share birth month	Gait assessment	Working memory	Info on time of the day
Can share birth year	Walking speed	Executive function	Info on fasting or no fasting
Can share calculated age	Oxygen uptake	Verbal memory	Info hours since last meal
Mortality general info	Lung function	Visual memory	Info on time to Scan/Assessment
Mortality Date of death	Hip circumference	Memory self-assessment/Everyday memory	<b>Hormones</b>
<b>Demographics</b>	Foot dominance	Verbal fluency	measured
<b>Language</b>	Body composition (fat, muscle)	Simple reaction time	Used test
Language of tests	Smell recognition	Processing speed	Acquired marker
Mother tongue/native	<b>Physical Health</b>	Motor speed/function	Metabolites
Bilingual	ADL Activities of Daily Living	Cognitive screening test / global cognition	measured
Second language	Alcohol	<b>Cognitive test miscellaneous</b>	Used test
<b>Education</b>	Smoking	<b>MRI</b>	Acquired markers
Subject education	Substance use	Scanner vendor	Immune marker
Mother education	Physical activity	Scanner type	measured
Father education	Medication	Field strength	Used test
Guardian education	Sleep	Number of channels Headcoil	Acquired markers
Partner education	Diet	T1-weighted	Miscellaneous
<b>Occupation</b>	<b>Mental health</b>	Sequence name	Vitamin D
Subject /Partner/Parent/gardian	Attachment style	Orientation	Cholesterol
	Anxiety	voxel dimensions	Retinol
	Worry	TR in seconds	Fatty acids
	Depression/mood	Number of slices	<b>DNA</b>
	Phobias	duration	Blood
<b>Housing</b>	Sleep problems/fatigue	T2-weighted	Genetic info available
Info on type of housing	Somatic symptoms	Sequence name	Whole -genome
Info on geographic location	Childhood trauma	orientation	# SNPs
<b>Income</b>	Diagnosis	voxel dimensions	type of analysis /chip
Household income (per year,...)	Stress	TR in seconds	Custom genotyping
Person (per year, ...)	Life events/experiences	Number of slices	type of analysis
Partner (per year, ...)	Social contact	duration	Measured polymorphisms/Genes
Parents/Guardian (per year, ...)	Retirement	DWI	Saliva
Currency	Quality of life	Sequence name	Genetic info available
<b>Marital status/relation</b>	Personality	number of directions	Whole -genome
<b>Family composition</b>	MMSE Mini Mental State Examination	Number of b-values used	# SNPs
<b>Social Network</b>	Psychiatric Family history	b-values	type of analysis
	Behavioral disturbances	TR in seconds	Custom genotyping
		Number of volumes / b-value	type of analysis
		duration	Measured polymorphisms/Genes
		rs-fMRI	
		Sequence name	
		Instruction: eyes open/ closed / fixation	
		TR in seconds	
		Number of volumes	
		Duration	
		Task fmri	
		paradigm 1	
		paradigm 2	
		paradigm 3	
		Miscellaneous	
		ASL	
		MRS	
		FLAIR	
		T2*	
		MT/noMT	

FIGURE 2. LIFE BRAIN DATA DIMENSIONS

## 1.2. Selection of variables of initial focus

Based on the detailed overview of in principle shareable data it has been decided, which variables to initial focus on (see Figure 3). The decision was informed by observed availability of data across sites, the focus of the research in WP4, specifically Task 4.2 “Analysis of individual pathways/mediator variables in relation to risk for mental health problems and resilience” (M18-M48), and Task 4.3 “Analysis of individual pathways/mediator variables in relation to risk for poor cognitive function and decline versus resilience” (M18-M 48).

The decision was to focus our efforts on the dimensions:

- Study set up/participants
- Demographics
- Physical Measures
- Physical Health
- Mental Health and
- Cognitive Tests.

The level of detail obtained for the MRI, Biomarkers, and DNA data dimensions was sufficient to inform the specific tasks on these dimensions i.e.

- Task 3.2. “Development and validation of a posteriori harmonization of MRI data” (M1-M9);
- Task 2.4 “Enrichment of existing cohorts by home collection of dried blood spots (DBS)” (M6–M36);
- Task 2.5 “Enrichment of existing cohorts by administration of buccal swabs for epigenetic analysis”(M6–M36); and
- Task 2.6 “Generation of new GWAS (genome-wide association study) data and harmonization of genetic data across samples” (M6-M24).

Study set up/ participants	Physical Health
<b>Setup</b>	Alcohol
Start and end year	Smoking
Time between subsequent assessments	Substance use
Assessment date	Physical activity
<b>Participants</b>	Medication
Number of subjects	Sleep
<b>Sex</b>	<b>Mental health</b>
Number of Females	Anxiety
Number of Males	Depression/mood
<b>Age related</b>	Stress
Age Range	Life events/experiences
Can share birth date	Personality
Can share birth month	MMSE Mini Mental State Examination
Can share birth year	<b>Medical Health</b>
Can share calculated age	physical Examination
Mortality general info	info on general health
Mortality Date of death	info on cardio vascular disease
<b>Demographics</b>	info on Vascular risk factors
<b>Language</b>	info on stroke
Language of tests	info on diabetes
<b>Education</b>	info on Longstanding illness
Subject education	info on Sickness absence
Mother education	<b>Cognitive Tests</b>
Father education	IQ or equivalent
Guardian education	Working memory
<b>Income</b>	Verbal memory
Household income (per year,...)	Visual memory
Currency	Memory self-assessment/Everyday memory
<b>Marital status/relation</b>	Verbal fluency
<b>Physical measures</b>	Processing speed
Weight	<b>MRI</b>
Weight at birth	<b>Biomarkers</b>
Height	<b>DNA</b>
Grip strength	
Waist circumference	
Blood pressure	
Handedness	

FIGURE 3. DIMENSIONS AND VARIABLES OF INITIAL FOCUS

Data harmonization of the MRI, Biomarker and DNA data is closely coupled to the work in these tasks.

### 1.3. Gathering study protocols and site-specific variables meta data

Next, spreadsheets were developed for gathering detailed information on respectively study protocols, i.e. employed inclusion and exclusion criteria and the order in which biological, physical, behavioural and cognitive, and neuroimaging assessments were performed (see Figure 4), and variable meta-data i.e. site-specific information on variable name, variable description, allowable values, data type, data format, allowable values, value description, and coding of missing data (see Figure 5).

Study:	HUBU																		
Timepoints	1	Instructions	Comments	2	Instructions	Comments	3	Instructions	Comments	4	Instructions	Comments	5	Instructions	Comments	6	Instructions	Comments	7
List order in which assessments were typically performed. Use headings for specific block of tests. Add /delete rows if necessary. The first column is just an example which you can overwrite. State test instructions in adjacent column if applicable																			
1	Child tests and questionnaires	Inventory	Questionnaire	1	Child tests and questionnaires	Inventory	Questionnaire	1	Child tests and questionnaires	Inventory	Questionnaire	1	Child tests and questionnaires	Inventory	Questionnaire	1	Child tests and questionnaires	Inventory	Questionnaire
2	MOT (CANTAB)	Followed Cantab instructions	Test	2	MOT (CANTAB)	Followed Cantab instructions	Test	2	MOT (CANTAB)	Followed Cantab instructions	Test	2	MOT (CANTAB)	Followed Cantab instructions	Test	2	MOT (CANTAB)	Followed Cantab instructions	Test
3	DMS (CANTAB)	Followed Cantab instructions	Test	3	DMS (CANTAB)	Followed Cantab instructions	Test	3	DMS (CANTAB)	Followed Cantab instructions	Test	3	DMS (CANTAB)	Followed Cantab instructions	Test	3	DMS (CANTAB)	Followed Cantab instructions	Test
4	RVP (CANTAB)	Followed Cantab instructions	Test	4	RVP (CANTAB)	Followed Cantab instructions	Test	4	RVP (CANTAB)	Followed Cantab instructions	Test	4	RVP (CANTAB)	Followed Cantab instructions	Test	4	RVP (CANTAB)	Followed Cantab instructions	Test
5	SST (CANTAB)	Followed Cantab instructions	Test	5	SST (CANTAB)	Followed Cantab instructions	Test	5	SST (CANTAB)	Followed Cantab instructions	Test	5	SST (CANTAB)	Followed Cantab instructions	Test	5	SST (CANTAB)	Followed Cantab instructions	Test
6	Break			6	Break			6	Break			6	Break			6	Break		
7	Verbal fluency	Animal and words (S) - 60 s	Test	7	Verbal fluency	Animal and words (S) - 60 s	Test	7	Verbal fluency	Animal and words (S) - 60 s	Test	7	Verbal fluency	Animal and words (S) - 60 s	Test	7	Verbal fluency	Animal and words (S) - 60 s	Test
8	RTI (CANTAB)	Followed Cantab instructions	Test	8	RTI (CANTAB)	Followed Cantab instructions	Test	8	RTI (CANTAB)	Followed Cantab instructions	Test	8	RTI (CANTAB)	Followed Cantab instructions	Test	8	RTI (CANTAB)	Followed Cantab instructions	Test
9	SOC (CANTAB)	Followed Cantab instructions	Test, half of subjects	9	SOC (CANTAB)	Followed Cantab instructions	Test, half of subjects	9	SOC (CANTAB)	Followed Cantab instructions	Test, half of subjects	9	SOC (CANTAB)	Followed Cantab instructions	Test, half of subjects	9	SOC (CANTAB)	Followed Cantab instructions	Test, half of subjects
10	RTI (CANTAB)	Followed Cantab instructions	Test	10	RTI (CANTAB)	Followed Cantab instructions	Test	10	RTI (CANTAB)	Followed Cantab instructions	Test	10	RTI (CANTAB)	Followed Cantab instructions	Test	10	RTI (CANTAB)	Followed Cantab instructions	Test
11	SWM (CANTAB)	Followed Cantab instructions	Test	11	SWM (CANTAB)	Followed Cantab instructions	Test	11	SWM (CANTAB)	Followed Cantab instructions	Test	11	SWM (CANTAB)	Followed Cantab instructions	Test	11	SWM (CANTAB)	Followed Cantab instructions	Test
12	FMC		Test, half of subjects	12	FMC		Test, half of subjects	12	FMC		Test, half of subjects	12	FMC		Test, half of subjects	12	FMC		Test, half of subjects
13				13				13				13				13			
14				14				14				14				14			
15				15				15				15				15			
16	Parent questionnaires			16	Parent questionnaires			16	Parent questionnaires			16	Parent questionnaires			16	Parent questionnaires		
17	Five-to-fifteen	Written on questionnaire	Questionnaire	17	BRIEF	Written on questionnaire	Questionnaire	17		None collected		17		None collected		17		None collected	
18	Screening	Written on questionnaire	Questionnaire	18	SDQ	Written on questionnaire	Questionnaire	18				18				18			
19				19	Demographic	Written on questionnaire	Questionnaire	19				19				19			
20				20				20				20				20			
21				21				21				21				21			
22				22				22				22				22			
23	MRI			23	MRI			23	MRI			23	MRI			23	MRI		
24	localizer			24	localizer			24	localizer			24	localizer			24	localizer		
25	T1 1			25	T1 1			25	T1 1			25	T1 1			25	T1 1		
26	T1 2			26	DWI			26	DWI			26	DWI			26	DWI		
27	T2_tse3d_vfl_ms_sag			27	DWI fieldmap 1			27	DWI fieldmap 1			27	DWI fieldmap 1			27	DWI fieldmap 1		
28	T2_pd_tse_tra			28	DWI fieldmap 2			28	DWI fieldmap 2			28	DWI fieldmap 2			28	DWI fieldmap 2		
29	DWI			29	T1 2			29	T1 2			29	T1 2			29	T1 2		
30	DWI fieldmap 1			30	T2_tse3d_vfl_ms_sag			30	T2_tse3d_vfl_ms_sag			30	T2_tse3d_vfl_ms_sag			30	T2_tse3d_vfl_ms_sag		
31	DWI fieldmap 2			31	T2_pd_tse_tra			31	T2_pd_tse_tra			31	T2_pd_tse_tra			31	T2_pd_tse_tra		
32				32	ep2d_diffsteaam_b1_BCO			32	ep2d_diffsteaam_b1_BCO			32	ep2d_diffsteaam_b1_BCO			32	ep2d_diffsteaam_b1_BCO		
33	Saliva samples			33	Saliva samples			33	Saliva samples			33	Saliva samples			33	Saliva samples		
34	Child tests and questionnaires			34	Child tests and questionnaires			34	Child tests and questionnaires			34	Child tests and questionnaires			34	Child tests and questionnaires		
35	Inventory	Followed Cantab instructions	Test	35	Inventory	Followed Cantab instructions	Test	35	Inventory	Followed Cantab instructions	Test	35	Inventory	Followed Cantab instructions	Test	35	Inventory	Followed Cantab instructions	Test
36	MOT (CANTAB)	Followed Cantab instructions	Test	36	MOT (CANTAB)	Followed Cantab instructions	Test	36	MOT (CANTAB)	Followed Cantab instructions	Test	36	MOT (CANTAB)	Followed Cantab instructions	Test	36	MOT (CANTAB)	Followed Cantab instructions	Test
37	DMS (CANTAB)	Followed Cantab instructions	Test	37	DMS (CANTAB)	Followed Cantab instructions	Test	37	DMS (CANTAB)	Followed Cantab instructions	Test	37	DMS (CANTAB)	Followed Cantab instructions	Test	37	DMS (CANTAB)	Followed Cantab instructions	Test

FIGURE 4. EXAMPLE SPREADSHEET LAYOUT FOR GATHERING SITE SPECIFIC STUDY PROTOCOLS



A	B	C	D	E	F	G	H	I	J	K	L
<b>PLEASE NOTE</b>	the goal is to get an overview how data is <b>CURRENTLY CODED</b> in the separate studies and assessment rounds	below are only some examples /snippets of some possible Tests (you can just deleted the rows)	Create Blue header rows for each test (example below)	Create new TABS for each Assessment round of a specific study and name TAB accordingly (i.e. Study 1,2,... Code Book Round 1,2, 3,...) and copy the first 3 rows from the Code Book Template into it	Allowable values for category variables with more than two response possibilities (not including missing values) should be listed in the LOOKUP Tab (see LOOKUP_example tab for guidance)						
Site:	INSERT SITE NAME HERE	Study:	INSERT STUDY NAME							Define missing data codes used in the studies on the Readme tab. ONLY LIST exceptions of your rule here!	specifics on coding scheme, reference to articles, other ....
Test_ID	Test_name	Variable_name	Variable_name_in_test	Variable_description	Type	Format /length	Allowable values	Value description	Examples	Missing data	Comments
SUBJECT	NA										
		Subject_ID	NA	Specific study subject id start with a character	string	7			hubu001		
		DOB	NA	Date of Birth	Date	YYYY-MM-DD			1957-12-02		
		DB	NA	Day of Birth	Integer	DD	1 - 31		2		
		MB	NA	Month of Birth	Integer	MM	1 - 12		12		
		YB	NA	Year of Birth	Integer	YYYY			1957		
		Sex	NA	biological sex	Categories	Enumeration	male, female		male		
DEMO	Demographics										
		education	NA	highest education attained	Categories	Enumeration	lookup table				
		edu_yrs		years of education	Integer	15	0-80				calculation method used
DMS	CANTAB: Delayed matching to										
		Investigator_ID	NA	investigator doing the testing	Categories	Enumeration	Lookup table		I001		
		DMS_time	NA	Test start time	Date	HH:MM:SS			12:12:56		
		DMS_date	NA	Test start date	Time	YYYY-MM-DD			1957-07-05		
		DMS_CL	NA	DMS Mean correct latency	Decimal	#.#			123,76655		
		DMS_CL_all	NA	DMS Mean correct latency (all delays)	Decimal	4,2			1233,76		
		DMS_CL_sim	NA	DMS Mean correct latency (simultaneous)	Decimal	#.#			123,76655		

**FIGURE 5. EXAMPLE SPREADSHEET FOR GATHERING SITE SPECIFIC VARIABLE META DATA**

## 1.4. Checking code book consistency

All sites provided the required information in study inclusion/ exclusion criteria and study protocols. Organization of the information on the specific instructions subjects were given during cognitive tests, part of the study protocols, is ongoing.

All sites provided information on variable metadata for the variables of initial focus. A stringent quality control of the site specific variable code books on consistency and completeness, with respect to reported information in the general shared variables overview, organization and formatting is ongoing.

The provided variable data allowed us to assess data availability as well the level of detail in which dimensions and variables of interest were assessed and sampled across Lifebrain sites. The latter allowed us to perform an overall evaluation of the data harmonization potential of the different types of variables.

## 1.5. Preliminary data harmonization potential

A detailed overview of the data available in Lifebrain is provided in Appendix 3.1- 3.3. In what follows, an overall evaluation of data harmonization potential of dimensions and cluster of variables is given.

1. Date of assessments, time of assessments, investigator doing the assessment, and subject sex, birth date, birth month, birth year, calculated age, and general info on mortality, and date of death.

This data does not need harmonization in the sense that scores, except for "Investigator doing the assessment" have the same meaning across sites. Sex will be recoded to female, male. Overall, for all variables that will be used in Lifebrain, we will impose a stringent coding regime i.e. organization and formatting of the meta data associated with Lifebrain variables. We strive to follow international standards and standards used in other large consortia comparable to Lifebrain in scope and focus, in order to enable efficient future coupling of Lifebrain data to that of other consortia.

2. Subject, mother, father, guardian, person accompanying child educational

Lifebrain has a good coverage of subject education. The available data will allow two strategies:

- a. Record highest attained education and convert to years of education following the International Standard Classification of Education (current version 2011) (3a, b).
- b. Re-categorize available data into approximately four categories (see for example, 3). Additionally, we will statistically account for possible cohort and country effects. While not available in all studies, importantly, information on parent education is available or available in principle in studies including partially (LCBC) or exclusively (HUBU, CALM) children. Of note, in CALM, education was registered for the person who accompanied the child. Education of mother/father/guardian/other can be retrieved from this data.

3. Subject occupation

Information is in principle available for most of the studies. If feasible, we will record and convert acquired data using International Standard Classification of Occupations (ISCO) 08 (4).

#### 4. Income

Data is not available in all studies. Harmonization of and/or including income in statistical models for future specific research questions is not straight forward and need to account for national differences in e.g. cost of daily living, housing prices, taxes, inflation etc (5).

#### 5. Marital and relationship status

Harmonization will include re-categorisation to ensure uniform coding.

#### 6. Weight, Weight at birth, Height, Grip strength, Waist circumference, Blood pressure, Handedness

Data harmonization on these variables will in principle only requires re-coding. However, we need to account for differences between sites in the protocols and methods used to derive these measures (e.g. Weight measured using a weight balance or reported; Blood pressure measured sitting, lying down, or both, ...).

#### 7. Alcohol

Most sites assessed alcohol use. However, assessments differ substantial in the level of detail of acquired information (e.g. typical/ heavy drinking, type of alcohol beverage, quantity, frequency, drinking pattern, reference period) and instruments used. Successful harmonization/ modelling of alcohol consumption will depend on which studies are included in specific investigations. Cultural and cohort differences need to be accounted for (6).

#### 8. Smoking

Most sites assessed smoking habits. However, assessments differ greatly in the level of detail of acquired information (type: cigarettes, cigars, pipe, quantity, frequency, reference time, age of starting/stopping). Categorizations such as "never", "current", "not current" or ever smoked "yes/no" might be possible. Depending on the studies included in specific investigations definition of variables on quantity /frequency may be feasible (7).

#### 9. Substance use

Substance use has been assessed in several studies. However, assessments differ in the level of detail (e.g. status, type of drug, frequency, age started/stopped). Depending on studies and the number of subjects that have used substances of abuse, we may be able to model its possible effects, for example using a categorical variable "never", "current", "not current" and/or exclude subjects from analyses.

## 10. Physical activity

Most sites assessed physical activity. However, the level of detail of acquired information (e.g. leisure, occupational, transport, daily living, organized, type, duration, regularity, intensity, frequency, MET or Calories equivalents) differs across sites as well as the assessment instruments that were used. We strive to harmonize on intensity, frequency, duration, and/or a categorical variable for level of activity (8,9).

## 11. Medication

Most sites assessed medication use. However, level of detail (e.g medication history, types of medication, medication codes, frequency, doses) and assessments instruments used differ greatly across sites. Data harmonization likely will involve categorical variables.

## 12. Sleep

Most sites assessed sleep. Eight studies used the Pittsburgh Sleep Quality Index. Others used in-house questionnaires. The data harmonization potential is deemed good and allows identifying common variables.

## 13. Anxiety, Depression and Mood

Most studies assessed anxiety and depression and mood variables. However, there is considerable variation in the specific assessment instruments that were used as well as the number of instruments used in each individual study. Moreover, not all studies can provide item level data. We will identify common variables, and assess the feasibility of implementing e.g. latent factor models, item response theory (IRT) (9, 10).

## 14. Memory self-assessment and Everyday memory, Stress, Life events and experience, Personality

These dimensions were assessed in several studies. However, there is considerable variation in the specific assessment instruments that were used as well as the number of instruments used in each individual study. Moreover, not all studies can provide item level data. We will identify common variables, and assess the feasibility of implementing e.g. latent factor models, item response theory (IRT) (9, 10, 11).

## 15. Physical Examination

Physical examinations were performed in several studies. However, level of detail and assessments instruments used differ greatly across sites. Information on general health, cardio vascular disease, diabetes, longstanding illness, stroke, vascular risk factors or sickness absence seems available. We strive to identify common variables and explore the possibility of creating categorical variables.

## 16. IQ or equivalent, Working memory, Verbal and visual memory, Mini Mental State Examination, Verbal fluency

Most studies assessed these variables. However, studies differ by the specific tests that were employed as well as by the number of test used in each individual study. We will identify common variables, and assess the feasibility of implementing e.g. latent factor models (10, 13, 14). The Mini Mental State Examination can be readily harmonized. Assessed letter and/or semantic fluency, while differing in used protocols, will likely allow us to normalized scores for this variable. Most sites assessed letter and/or semantic fluency allowing us to generate common variables and possibly use normalized scores.

## 17. Processing speed

Several sites assessed processing speed. However, studies differ by the specific tests that were employed as well as by the number of test used in each individual study. We will identify common variables, and assess the feasibility of implementing e.g. latent factor models (10, 13, 14). Of note, most studies employed versions of the trail making task, making it feasible to use normalized scores.

## 2. Next steps

We continue with quality control and completion of site-specific study protocols. The latter specifically concerns documenting the instructions given to subjects during cognitive tests. Additionally, we continue with stringent quality control of the site-specific variable code books on completeness, organization and formatting.

Lifebrain will allow for different approaches depending on the specific research questions and hypotheses. It will allow for centralized analyses by creating one dataset for analyses, which requires harmonization of targeted datasets, or coordinated independent analyses (1).

## 2.1. General harmonization procedures

The current deliverable gathered and organized information about all variables of interest across cohorts in Lifebrain. While harmonization of some of the dimensions/variables will be straightforward, other dimensions/variables will require more effort to harmonise given the large variation in assessment instruments used for some of the dimensions. In our data harmonization efforts we will build on other initiatives, such as EPOSA (16), and BIOSHARE (17) (see Figure 6).

In the autumn of 2017, we will establish specific workgroups for data harmonization of specific dimensions/variables. The task groups will generate guidelines on how to harmonize specific dimensions/variables, and on what analytic approaches are feasible and/or recommendable. It is important to determine whether each of the dimensions/variables measures the same concept in order for the harmonization to proceed. If the variable does not measure the same concept, harmonization of the data will not be possible. If the variable does measure the same concept, then the harmonization procedures can proceed.

Another important step is to determine whether response categories overlap. If they do not overlap, variables will be matched on categories, e.g. smoking yes/no. If they do overlap, the task group will determine possible harmonization algorithms that allows for as little loss of information as possible.

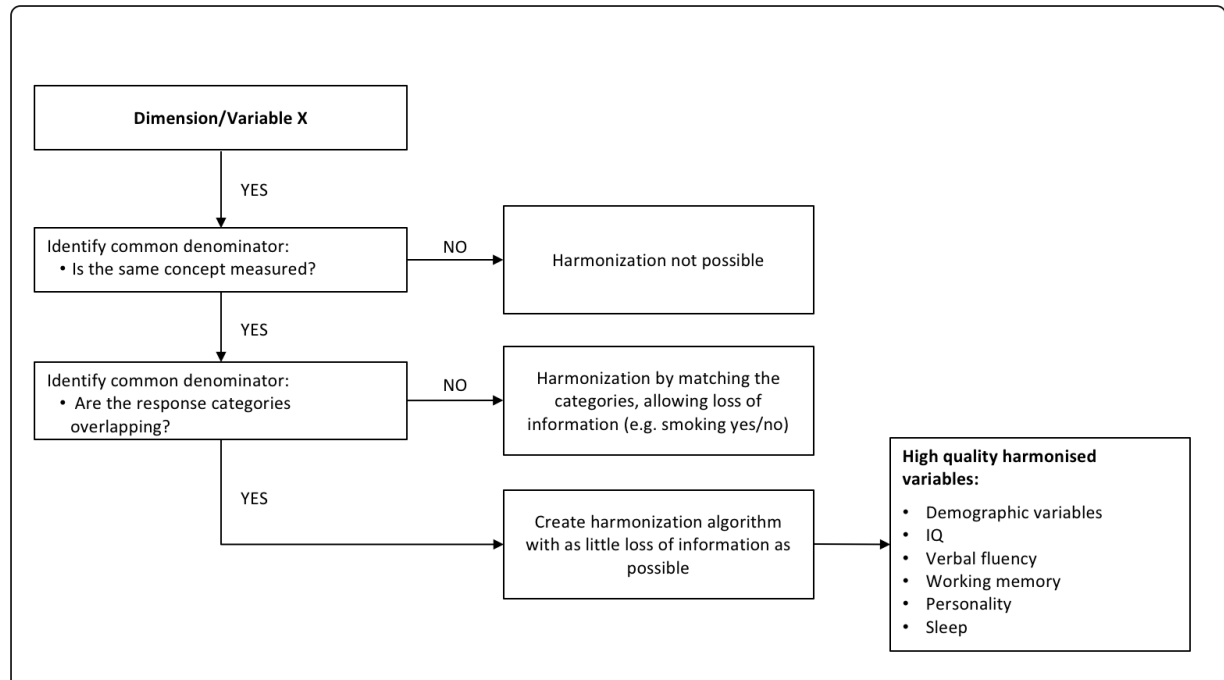


FIGURE 6. SCHEMATIC PRESENTATION OF THE HARMONIZATION PROCESS FOR ALL DIMENSIONS/VARIABLES ADAPTED FROM (16)

## 2.2. Designing the Lifebrain database

In the autumn of 2017 and in parallel with the data harmonization efforts we will start with designing the Lifebrain database tables that will ultimately populate the Lifebrain database using the infrastructure developed in Task 3.1 “Development of a data storage and management system” (M1-M6), and develop conversion scripts to populate Lifebrain tables with site-specific data.

## 3. Conclusion

The objectives of the current deliverable were fulfilled as detailed below:

- Map and categorize available data for all Lifebrain studies and time points: Completed
- Identify an initial set of variables of interest: Completed
- Obtain site specific code books, which contain information on study specific designs, standard operating procedures, data collection methods, and metadata on variables e.g. variable names, variable description, data type, data format, allowable values etc.
  - Site-specific code books, which contain information on study specific designs, standard operating procedures were generated. Gathering the instructions given to subjects during test is in progress.
  - All sites provided information on variable metadata. The detailed quality control of the site specific variable code books on consistency and completeness, with respect to reported information in the general shared variables overview, and formatting is still in progress.
- Evaluate commonalities across sites and overall harmonization potential
  - Preliminary data harmonization potential of initial variables of interest has been evaluated

## 4. References

- (1) Hofer SM, Piccinin AM. Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychol Methods*. 2009;14(2):150-164.
- (2) Maelstrom Research guidelines for rigorous retrospective data harmonization. Fortier I, Raina P, Van den Heuvel ER, Griffith LE, Craig C, Saliba M, Doiron D, Stolk RP, Knoppers BM, Ferretti V, Granda P, Burton P. *Int J Epidemiol*. 2017 Feb 1;46(1):103-105.

- (3a) [http://ec.europa.eu/eurostat/statistics-Explained/index.php/International\\_Standard\\_Classification\\_of\\_Education\\_\(ISCED\)](http://ec.europa.eu/eurostat/statistics-Explained/index.php/International_Standard_Classification_of_Education_(ISCED))
- (3b) <http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-06-14-246>
- (4) <http://www.ilo.org/public/english/bureau/stat/isco/>
- (5) Canberra Group (2011) Handbook of Household Income Statistics. Second Edition 2011. Geneva: United Nations.
- (6) Agrawal A, Freedman ND, Cheng YC, Lin P, Shaffer JR, Sun Q, Taylor K, Yaspan B, Cole JW, Cornelis MC, DeSensi RS, Fitzpatrick A, Heiss G, Kang JH, O'Connell J, Bennett S, Bookman E, Bucholz KK, Caporaso N, Crout R, Dick DM, Edenberg HJ, Goate A, Hesselbrock V, Kittner S, Kramer J, Nurnberger JI, Jr., Qi L, Rice JP, Schuckit M, van Dam RM, Boerwinkle E, Hu F, Levy S, Marazita M, Mitchell BD, Pasquale LR, Bierut LJ. Measuring alcohol consumption for genomic meta-analyses of alcohol intake: opportunities and challenges. *Am J Clin Nutr*. 2012;95(3):539-547.
- (7) <http://www.dynamo-hia.eu/dsresource?type=pdf&disposition=inline&objectid=rivmp:230417&versionid=&subobjectname=>
- (8) Warren JM, Ekelund U, Besson H, Mezzani A, Geladas N, Vanhees L. Assessment of physical activity - a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *Eur J Cardiovasc Prev Rehabil*. 2010;17(2):127-139.
- (9) Pettee Gabriel KK, Morrow JR, Jr., Woolsey AL. Framework for physical activity as a complex and multidimensional behavior. *J Phys Act Health*. 2012;9 Suppl 1:S11-18.
- (10) McArdle JJ, Grimm KJ, Hamagami F, Bowles RP, Meredith W. Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychol Methods*. 2009;14(2):126-149.
- (11) van den Berg SM, de Moor MH, McGue M, Pettersson E, Terracciano A, Verweij KJ, Amin N, Derringer J, Esko T, van Grootheest G, Hansell NK, Huffman J, Konte B, Lahti J, Luciano M, Matteson LK, Viktorin A, Wouda J, Agrawal A, Allik J, Bierut L, Broms U, Campbell H, Smith GD, Eriksson JG, Ferrucci L, Franke B, Fox JP, de Geus EJ, Giegling I, Gow AJ, Grucza R, Hartmann AM, Heath AC, Heikkila K, Iacono WG, Janzing J, Jokela M, Kiemeny L, Lehtimäki T, Madden PA, Magnusson PK, Northstone K, Nütle T, Ouwens KG, Palotie A, Pattie A, Pesonen AK, Polasek O, Pulkkinen L, Pulkki-Raback L, Raitakari OT, Realo A, Rose RJ, Ruggiero



D, Seppala I, Slutske WS, Smyth DC, Sorice R, Starr JM, Sutin AR, Tanaka T, Verhagen J, Vermeulen S, Vuoksima E, Widen E, Willemsen G, Wright MJ, Zgaga L, Rujescu D, Metspalu A, Wilson JF, Ciullo M, Hayward C, Rudan I, Deary IJ, Raikonen K, Arias Vasquez A, Costa PT, Keltikangas-Jarvinen L, van Duijn CM, Penninx BW, Krueger RF, Evans DM, Kaprio J, Pedersen NL, Martin NG, Boomsma DI. Harmonization of Neuroticism and Extraversion phenotypes across inventories and cohorts in the Genetics of Personality Consortium: an application of Item Response Theory. *Behav Genet.* 2014;44(4):295-313.

(12) Duzel S, Voelkle MC, Duzel E, Gerstorf D, Drewelies J, Steinhagen-Thiessen E, Demuth I, Lindenberger U. The Subjective Health Horizon Questionnaire (SHH-Q): Assessing Future Time Perspectives for Facets of an Active Lifestyle. *Gerontology.* 2016;62(3):345-353.

(13) Brandmaier AM, von Oertzen T, McArdle JJ, Lindenberger U. Structural equation model trees. *Psychol Methods.* 2013;18(1):71-86.

(14) Bender AR, Prindle JJ, Brandmaier AM, Raz N. White matter and memory in healthy adults: Coupled changes over two years. *Neuroimage.* 2016; 131:193-204.

(15) Costa A, Bak T, Caffarra P, Caltagirone C, Ceccaldi M, Collette F, Crutch S, Della Sala S, Demonet JF, Dubois B, Duzel E, Nestor P, Papageorgiou SG, Salmon E, Sikkes S, Tiraboschi P, van der Flier WM, Visser PJ, Cappa SF. The need for harmonisation and innovation of neuropsychological assessment in neurodegenerative dementias in Europe: consensus document of the Joint Program for Neurodegenerative Diseases Working Group. *Alzheimers Res Ther.* 2017;9(1):27.

(15) Sachdev PS, Lipnicki DM, Kochan NA, Crawford JD, Rockwood K, Xiao S, Li J, Li X, Brayne C, Matthews FE, Stephan BC, Lipton RB, Katz MJ, Ritchie K, Carriere I, Ancelin ML, Seshadri S, Au R, Beiser AS, Lam LC, Wong CH, Fung AW, Kim KW, Han JW, Kim TH, Petersen RC, Roberts RO, Mielke MM, Ganguli M, Dodge HH, Hughes T, Anstey KJ, Cherbuin N, Butterworth P, Ng TP, Gao Q, Reppermund S, Brodaty H, Meguro K, Schupf N, Manly J, Stern Y, Lobo A, Lopez-Anton R, Santabarbara J. COSMIC (Cohort Studies of of Memory in an International Consortium): an international consortium to identify risk and protective factors and biomarkers of cognitive ageing and dementia in diverse ethnic and sociocultural groups. *BMC Neurol.* 2013;13:165.

(16) Schaap LA, Peeters GM, Dennison EM, Zambon S, Nikolaus T, Sanchez-Martinez M, Musacchio E, van Schoor NM, Deeg DJ. European Project on OsteoArthritis (EPOSA): methodological challenges in harmonization of existing data from five European population-based cohorts on aging. *BMC Musculoskelet Disord.* 2011;12:272.

(17) Doiron D, Burton P, Marcon Y, Gaye A, Wolffenbuttel BH, Perola M, Stolk RP, Foco L, Minelli C, Waldenberger M, Holle R, Kvaloy K, Hillege HL, Tasse AM, Ferretti V, Fortier I. Data harmonization and federated analysis of population-based studies: the BioSHaRE project. *Emerg Themes Epidemiol.* 2013;10(1):12.

## 5. Appendix

Appendix 1: Abbreviations of test and questionnaires that have been employed in Lifebrain studies (PDF)

Appendix 2: Number of subjects and observations (PDF)

Appendix 3.1: Detailed overview of the data available in Lifebrain (PDF)

Appendix 3.2: Detailed overview of the data available in Lifebrain (PDF)

Appendix 3.3: Detailed overview of the data available in Lifebrain (PDF)

Abbreviations	Full name
(Child and Parents)	Indicates that both the participant and parents have filled out a version of the questionnaire
(Parents)	Indicates that the parents filled out the questionnaire
4D5Q, 4DKL (in German)	The Four-Dimensional Symptom Questionnaire
AC	Alcohol Consumption
ACE-R	Addenbrookes Cognitive Examination Revised
AGORA	Angoraphobia
AH4	non-verbal reasoning part of the Alice Heim Group Ability Test
ANT, Emotional ANT	Attention Networks Test
ASEBA	Achenbach System of Empirically Based Assessment
ASI	Addiction Severity Index
AUDIT	Alcohol Use Disorders Identification Test
BAECKE	Baecke Physical Activity Questionnaire
BAI	Beck Anxiety Inventory
BASE activity questionnaire	Berlin Ageing Study activity questionnaire
Basecog	Berlin Ageing Study Cognitive test battery
BDI	Beck Depression Inventory
BIG five	The five-factor model of personality
BIS-11	Barrett's Impulsiveness Scale
BIS/BAS	Behavioural Inhibition/Approach System
BNT-60	Boston Naming Test
BPRS	Brief Psychiatric Rating Scale
BRIEF	Behaviour Rating Inventory of Executive Function
Brugha	aka The List of Threatening Experiences (LTE)
CAGE	Acronym created from 4 alcohol dependence questions (Ewing, 1984)
CAMB	Copenhagen Aging and Midlife Biobank
CAMT	Cimbi affective memory task
CASE	Child and Adolescent Survey of Experiences
CASP	control', 'autonomy', 'pleasure' and 'self-realization'; quality of life measure in older age
CC3000	Cambridge Centre for Ageing Neuroscience 3000 sample
CDDR	Customary Drinking and Drug Use Record
CDQ	Cognitive Dysfunction questionnaire
CERAD	The Consortium to Establish a Registry for Alzheimer's Disease
CES-D	Centre for Epidemiological Studies Depression Scale
CHAMPS	CHAMPS Physical Activity Questionnaire for Older Adults
CIDI	Composite International Diagnostic Interview (WHO)
CIS-R	Clinical Interview Schedule – Revised
CLOX	Clock drawing task
CN / 10MQ	10 memory self-report questions
CPT	Continuous Performance Test
CVA	CerebroVascular Accident
CVLT	California Verbal Learning Test
DAST-20	Drug Abuse Screening Test
DE	Dependence Symptoms
DHEA-5	Dehydroepiandrosterone 5 mg
DKEFS	Delis-Kaplan Executive Function System
DMS(Cantab)	Delayed Matching to Sample
DSM-III-R	Diagnostic and Statistical Manual of Mental Disorders III revised
E2	Estradiol
EFIT	Emotional face identification task
EH1	Edinburgh handedness inventory
EMO	Emotional go/nogo task
EMQ	Everyday memory questionnaire (Sunderland et al. 1983)
EPI	Echo-Planar Imaging
EPIC	European Prospective Investigation into Cancer and Nutrition
EPIC EPAQ-2	EPIC Physical Activity Questionnaire
ERQ	Emotion Regulation Questionnaire
F-HAM	Family history assessment module
FAB	Frontal Assessment Battery
FCSRT	Free and Cued Selective Reminding Test (Buschke)
FEAR	FEAR Questionnaire
FFQ	Food Frequency Questionnaire
Framingham CHD	Framingham Coronary Heart Disease risk score
Framingham CVD	Framingham cardiovascular disease risk score
FTF	Five to fifteen
FTND	the Fagerström Test for Nicotine Dependence
GAD	Generalized Anxiety Disorder
GDS	Geriatric Depression Scale (Yesavage)
GHQ	General Health Questionnaire
GMS -A/B	Geriatric Mental State (GMS), Version A/B (Copeland et al.,1988)
HADS	Hospital Anxiety and Depressive Scale
HAMD	Hamilton Depression Scale
HAS items	History and Aetiology Schedule
HC	Head coil
HDRS	Hamilton Depression Rating Scale
HiPIC	Hierarchical Personality Inventory for Children
HNC	Head-neck coil
HSP	Highly Sensitive Person
HU	Harmful alcohol Use
HVLT-R	Hopkins Verbal Learning Test-Revised
IADL	Instrumental Activities of daily Living (Lawton & Brodsky 1969)
ICU	Inventory of Callous-Unemotional traits
IDS	Inventory of depressive symptomatology
IPAQ	International Physical Activity Questionnaire
IPAQ-S	International Physical Activity Questionnaire short version 2002
IST-2000-R	Intelligens Struktur Test 2000 Revised
J-EPQ	Junior Eysenck personality questionnaire
Jenkins	Jenkins Sleep Questionnaire
JLO	Judgement of Line Orientation
KSQ	Karolinska Sleep Questionnaire
LOT-R	Life Orientation-Revised
LTE-Q	List of Threatening Experiences questionnaire
MACA	MacArthur stress reactivity questionnaire
MADRS	Montgomery-Asberg Depression Rating Scale
MASQ	Mood and Anxiety Symptom Questionnaire
MB	Multi-band
MDI	Major depression inventory
MDQ	Mood Disorder Questionnaire
MEMPR	Multi-Echo MPAGE
MET	Metabolic Equivalent Total
MFQ	Memory Functioning Questionnaire (MFQ) (Gilewski 1990)
MID	Major depression inventory
MIDI	Midlife Development Inventory (MIDI) Scale
Mill Hill	Mill Hill vocabulary scale
MMSE	Mini Mental State Examination
MNA	Mini Nutritional Assessment
MOCA	Montreal Cognitive Assessment
MPRAGE	Magnetization-Prepared Rapid Gradient Echo
MTCF	Modified Taylor Complex Figure Test
N-back	Working Memory Test
NA (VUMC)	Negative Affect
NART	National Adult Reading Test
Neo- Pi-R	NEO Personality Inventory-Revised
NEO-FFI	Neuroticism Extraversion Openness Five-Factor Inventory
NEPSY	a neuropsychological assessment instrument
NEPSY VA	NEPSY- Visual Attention subtest
NPI	Neuropsychiatric Inventory (Cummings)
PA	Panic Disorder
PA (VUMC)	Positive Affect
PAL	Paired Associated Learning (visual memory and learning) CANTAB
PASE	Physical Activity Scale for the Elderly
PGCMS	The Philadelphia Geriatric Center (PGC) Morale Scale
PHQ-9	Perceived Health questionnaire
PING	Pediatric Imaging, Neurocognition, and Genetics
PSQI	Pittsburgh Sleep Quality Index
PSS	Perceived Stress Scale
PSWQ	Penn-State Worry Questionnaire ultrabrief version
RAVLT	Rey Auditory Verbal Learning Test
RCFT	Rey Complex Figure Test
Rey-0	Rey–Osterrieth complex figure test
ROCF	Rey-Osterreich complex figure
RQ	The Relationship Questionnaire (Bartholomew & Horowitz, 1991)
RTI (CANTAB)	Cambridge Neuropsychological Test Automated Battery Reaction Time touchscreen version
RVP	Rapid Visual Information Processing (CANTAB)
SA (VUMC)	Somatic Arousal
SCD-Q	Subjective Cognitive Decline Questionnaire (Rami et al., 2014)
SCID-I	Structured Clinical Interview for DSM-IV-TR Axis I Disorders
SCL-90	Symptom Checklist 90 items
SDMT	Symbol Digit Modalities Test
SDQ	Strength and difficulties questionnaire
SF-36	36-Item Short Form Health Survey
SHBG	Sex Hormone Binding Globulin
SOC (CANTAB)	Stockings of Cambridge
SOEP	German Socio-Economic Panel Study (SOEP)
SPSRQ-CR	Sensitivity to Punishment and Sensitivity to Reward Questionnaire - Child Revised
SST (CANTAB)	Stop-Signal Task
STAI	State and Trait Anxiety Inventory
STAI-C	State Trait Anxiety Inventory for Children
STW	Spot the word
SWLS	The Satisfaction With Life Scale
SWM (CANTAB)	Spatial Working Memory
TCF	Taylor Complex Figure Test
TCI	Temperament and Character Inventory
TFEQ	Three Factor Eating Questionnaire
TMT A,B	Trail Making Test A,B
TMT-A	Trail Making Test part A
TMT-B	Trail Making Test part B
TOPF	Test of Premorbid Functioning
TOT	Tip of The Tongue
UCLA	UCLA University of California, Los Angeles) Loneliness Scale
UPSIT	University of Pennsylvania Smell Identification Test
VOSP	Visual Object and Space Perception Battery
WAIS	Wechsler Adult Intelligence Scale
WAIS-IV DS, DC	Digit Span (DS) and Coding (DC) tests from the Wechsler Adult Intelligence Scale - Fourth Edition (WAIS- IV;
WASI	Wechsler Abbreviated Scale of Intelligence
WHIS	Women's Health Initiative Insomnia Rating Scale
WHODAS	World Health Organization Disability Assessment Schedule
WISC	Wechsler Intelligence Scale
WMS III	Wechsler Memory Scale-3 <sup>rd</sup> ed
WPPSI	Wechsler Preschool and Primary Scale of Intelligence
WQ	Women's questionnaire
Young	Young Mania Rating Scale

Appendix 2

Lifebrain: overview of the total number of unique participants and observations

Lifebrain	Site	UiO	UmU	UOXF	MPIB		UB						REGIONH		VUMC	MRC	UCAM
All studies	Study	LCBC	Betula	Whitehall II Imaging	BASE	BASE II	WAHA	MSA_PSP	PD	GABA	iTBS	CR	HUBU	LISA	NESDA	CamCAN	CALM
	Number of time points	3	3	1	8	2	3	1	3	1	1	1	12	1	6	3	1
12527	Total unique subjects	1677	376	800	516	2200	82	24	46	33	27	49	94	450	2981	2690	482
7327 7387	Behaviour only	39		25	516	1755								90	2680	2057	255
5200 5140	MRI+Behavior	1638	376	775		445	82	24	46	33	27	49	94	390 450	301	633	227
27972	Total observations	2533	707	800	1402	2527	229	24	105	33	27	49	817	450	14860	2927	482
19873	Behaviour only	39		25	1402	1755	52							60	14228	2057	255
8099	MRI+Behavior	2494	707	775		772	177	24	105	33	27	49	817	390	632	870	227







Detailed overview of the data available in Lifebrain																				
include	Categories	Variables	Subvar	Site																
1				UIO	UMU	UOXF	MPiB				UB					REGIOH		VUMC	MRC	UCAM
1	Study	Study	Study name	LCBC	Betula:	Whitehall II	BASE	BASE-II	WAHA	MSA_PSP	PD	GABA	ITBS	CR	HUBU	LISA	NESDA	CamCAN	CALM	
1	Study	Study	Time points	3	2	1	8	2	3	1	3	1	1	1	12	1	6	3	1	
0	Mental health	Attachment style	Measured	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	
0	Mental health	Attachment style	Child and parent Social dominance orientation + Parental Bonding Inventory, Interpersonal relationships	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
1	Mental health	Anxiety	Assessed	1	1	1	1	0	0	0	1	0	1	1	1	1	1	1	1	
1	Mental health	Anxiety	STAI-C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
1	Mental health	Anxiety	ASI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
1	Mental health	Anxiety	BAI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
1	Mental health	Anxiety	BPRS	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Anxiety	CAMB quest	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
1	Mental health	Anxiety	CIDI	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
1	Mental health	Anxiety	FEAR	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
1	Mental health	Anxiety	GHQ	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Anxiety	GMS-A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Anxiety	HADS	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	
1	Mental health	Anxiety	NPI	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
1	Mental health	Anxiety	Revised Children's Anxiety and Depression Scale-25 (Choripita et	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
1	Mental health	Anxiety	SCL-90	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
1	Mental health	Anxiety	SDQ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Anxiety	STAI	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Anxiety	Ever been told by a doctor? unstructured interview (self report)	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	Mental health	Worry	Measured	0	0	0	1	?	0	0	0	0	0	0	0	0	0	1	0	
0	Mental health	Worry	Penn State Worry Questionnaire	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
0	Mental health	Worry	PSWQ	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	Mental health	Worry	Revised Children's Anxiety and Depression Scale-25 , child self-report and parent report	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
1	Mental health	Depression/mood	Measured	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	
1	Mental health	Depression/mood	Apathy Scale (Sergio Starkstein)	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	
1	Mental health	Depression/mood	BDI	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
1	Mental health	Depression/mood	BSI	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
1	Mental health	Depression/mood	CAMB quest	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
1	Mental health	Depression/mood	CES-D	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Depression/mood	CIDI	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
1	Mental health	Depression/mood	CIS-R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Depression/mood	DSM-III-R	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Depression/mood	GDS	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
1	Mental health	Depression/mood	GHQ	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Depression/mood	HADS	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	
1	Mental health	Depression/mood	HAMD	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Depression/mood	HDRS	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Depression/mood	IDS	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
1	Mental health	Depression/mood	MADRS (fMRI sample only)	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
1	Mental health	Depression/mood	MASQ	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
1	Mental health	Depression/mood	MDI	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
1	Mental health	Depression/mood	MDQ	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	
1	Mental health	Depression/mood	PGCMS	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Depression/mood	PHQ-9	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Depression/mood	Revised Children's Anxiety and Depression Scale-25 , child self-report and parent report	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
1	Mental health	Depression/mood	SCID-I	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Depression/mood	SDQ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Depression/mood	Young	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Mental health	Depression/mood	in-house structured interview	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	



1	DNA	Saliva	Measured	1	1	0	0	1	0	0	0	0	0	0	1	0	1	1	1
---	-----	--------	----------	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---