This webinar will be recorded
Recordings can be made available for knowledge dissemination.

Questions or remarks?
Please raise your hand by clicking on the “raise your hand”-symbol or use the chat

Mute your microphone
Please mute your microphone to minimize background noise by clicking on the “mute”-symbol on the top of your screen.

What can we improve for the future?
We are happy to hear any additional questions or feedback for our next expert meeting. Please contact us by mail: robin.verjans@lygature.org
Virtual Expert Meeting Series:
How to Make Health(care) Data Available for Research?

26th of November 2020
Health-RI COVID-19 Data Support Programme

1. Connecting communities
2. Expertise, tools & services
3. Best practices

To support investigators and healthcare professionals with tools and services in their search for ways to overcome the pandemic and its health consequences. Facilitate and stimulate an integrated health data infrastructure.
Connecting research and health communities (UMC’s, VWS, FMS, STZ, NFU, SURF, RIVM)

COVID-19 collections overview

COVID-19 Collections

Collections encompassing data, samples, images or biobank services.

COVID-19 initiatives overview

COVID-19 initiatives

Working groups, resources, data tools and services, funding, guidelines, and requests for collaboration.

ZonMW funded ‘National Observational COVID-19 Data Portal’
National Observational COVID-19 Data Portal

I and R of FAIR

F and A of FAIR
COVIDpredict
Het coronavirus vraagt veel van de Nederlandse zorg. Met name de intensive cares staan onder druk. Het is van groot belang dat de zorg voor COVID-19 patiënten zo goed en doelmatig mogelijk geleverd wordt. Daarom is het belangrijk om zo snel en veel mogelijk te leren van de reeds opgenomen COVID-19 patiënten in de Nederlandse ziekenhuizen. Door data over deze patienten te verzamelen, kunnen we de ziekte beter begrijpen en helpen bepalen wie welke behandeling en zorg moet krijgen.

BBMRI.nl
A portal to request samples, images and data.

Search

MMI partners

BBMRI.nl
healthRI
27/11/2020
Expertise, tools & services

- **XNAT** imaging archive toolkit
- **Podium** a portal to request samples, images and data
- **Personal Health Train (PHT)**
- **ELSI service desk**, support with ethical, legal and social implications
Facilitating best practices

- Ensuring uniform data collection in the Netherlands: COVID-19 research tips
- Harmonizing best practices for data sharing procedures – a guide on data governance

- Generic Policy for Access to and Sharing of COVID-19 Data, based on common Data Sharing Agreements
- Providing a legal framework for data and sample collections – a national COVID-19 citizen control registry
- Preserving patient material and samples – a national COVID-19 biobank
Virtual Expert Meetings Series

Goal

- To create a **discussion platform** to discuss current hurdles, experiences and best practices how to make health(care) data available for scientific (re)use.

Program

|   | Thursday October 29\textsuperscript{th} 14.00-15.00 | CBS - Fatima El Messlaki & Ivo Gorissen  
NHR - Saskia Houterman |
|---|---------------------------------------------------|---------------------------------------------------------------------|
| 1 | Thursday November 12\textsuperscript{th} 14.00-15.00 | Durrer Center - Erik van Iperen & Wanda Hermans-Van Ast  
IKNL - Peter Prinsen & Vincent Ho |
| 2 | Thursday November 26\textsuperscript{th} 14.00-15.00 | LifeLines - Trynke de Jong  
DataHub – Pascal Suppers |
This webinar will be recorded
Recordings can be made available for knowledge dissemination.

Questions or remarks?
Please raise your hand by clicking on the “raise your hand”-symbol or use the chat.

Mute your microphone
Please mute your microphone to minimize background noise by clicking on the “mute”-symbol on the top of your screen.

What can we improve for the future?
We are happy to hear any additional questions or feedback for our next expert meeting. Please contact us by mail: robin.verjans@lygature.org

The presentation will start soon

The presentation will start soon
Introducing the new LL dataplatform

BBMRI Expert Meeting, 26-11-2020

trynke de jong (t.r.de.jong@lifelines.nl)
to make data and samples from our longitudinal population cohort available worldwide for multidisciplinary research in the field of healthy ageing
167,000 people like you and me
- 3 Northern provinces (NL)
- general population (~10%)
- characteristics:
  - stable, do not move around
  - ethnically uniform
  - socio-economically diverse

all signed an Informed Consent form

our general assessments

2006
baseline assessment
questionnaire 1A
visit 1A1 - measurements
anthropometry, anxiety/depression (MINI), autofluorescence, blood pressure,
Ruff Figural Fluency Test (RFFT), Mini Mental State Examination (MMSE), ECG,
lung function test
visit 1A2 - biological samples
fasting blood, DNA, 24-hour urine, spot urine (morning)

2007
follow up questionnaire 1B

2008
follow up questionnaire 1C

2009
second assessment
questionnaire 2A
visit 2A1 - measurements
anthropometry, blood pressure, ECG, lung function test, Cogstate, Jump test, anxiety/depression (MIND)
visit 2A2 - biological samples
fasting blood, DNA, 24-hour urine, feces, scalp hair

2010
follow up questionnaire 2B

2011
third assessment
questionnaire 3A
visit 3A1 - measurements
anthropometry, blood pressure, ECG, lung function test, digital RFFT, MMSE, Cogstate, anxiety/depression (MINI)
visit 3A2 - measurements
fasting blood, 24-hour urine
anyone with a scientific track record & a good proposal (incl. companies but stricter rules)

fee-for-service non-profit

release in a secure workspace or HPC

return of sum scores, codes, sample results, add-on data to Lifelines

our data access procedures
additional assessments

open protocol -> additional data/sample collections:

- initiated by researchers
- from big & broad to small & precise
- use of various collection methods
- data is integrated with the existing database
- can be done rapidly if situation arises

(-> Covid-19)
N=167,000
Lifelines COHORT
(>10,000s ‘phenotypes’)

N=13,000
Lifelines "GWAS"
Illumina CytoSNP12

N=119
WGS/cell stim

N=1,500
Lifelines DEEP
Hum. Methyl. 450K

N=38,500
Lifelines UGLI
Illumina Global Screening Array

only partly integrated in database
- due to size
- linkage pseudonyms to data in HPC

incl. N=10,000: microbiome analyses

incl. 1,500 microbiome analyses
data linkage

Statistics Netherlands (CBS):
- government registries
- medication use
- cause of death

Medical Registries:
- GP’s
- PALGA
- IKNL

Not integrated in database:
- due to GDPR constraints
- dataset created on demand per project
- exception for linkage purely by location

Environmental data:
- noise
- air quality
- municipality data

LifeLines logo
how do I know *in advance* what I will get?
where can I find *variable x*?
how do I recognize *longitudinal* data?
what *metadata* can I expect?
who was *(not) invited* for a given study?
why is this particular cell *empty*?
when/how did the *protocol* change?
can I use prof. x’s *sum scores*?

how to *minimise* released data?
when to release *new* tables?
when do we *split/merge* tables?
solution = resolution
basic metadata per data point

**Participant-ID:** 123445  
**Sex:** female  
**DOB:** 01-01-1991  
**Postal code:** 9100  
**Subcohort:** UGLI, DAG3

**smoking_current_adu_q_1**  
“Do you currently smoke?”  
“Rookt u momenteel?”  
1 = Yes/Ja  
2 = No/Nee

**Date:** 11-12-2013  
**Assessment:** 2A  
**Element:** Questionnaire 1  
**Variant:** 18-64y, digital, v2

**Participant-ID:** 987643  
**Sex:** male  
**DOB:** 02-02-2002  
**Postal code:** 8770  
**Subcohort:** GWAS

**bodyweight_exact_chi_q_1**  
Bodyweight in kg (measured)  
Gewicht in kg (gemeten)  
Decimal

**Date:** 11-12-2010  
**Assessment:** 1A  
**Element:** Visit 1  
**Variant:** 8-17y, v1
**WHO**
- full list of participants (from CRM)
  - DOB, DOD
  - member of subcohort...
  - invited for element...
  - versions:
    - sex
    - postal code

**WHEN**
- full list of assessments, elements, variants

**WHAT**
- full list of variables
  - coded name
  - datatype, unit
  - label (Dutch/English)
  - enumerations (if any)
  - theme (section & subsection)

**WHO x WHEN**
- full list of participants per variant

**WHEN x WHAT**
- full list of variables per variant

**assessments**
- assessment = overarching data/sample collection with common goal

**element**
- element = part of an assessment that takes place on 1 day/date

**variant**
- variant = part of an element with a highly standardized protocol:
  - participant selection criteria
  - set of variables
  - collection method

Any changes? New variant!
new data catalogue

https://data-catalogue.lifelines.nl/

<table>
<thead>
<tr>
<th>Assessments</th>
<th>1a</th>
<th>1b</th>
<th>1c</th>
<th>2a</th>
<th>2b</th>
<th>sec</th>
<th>coqv</th>
</tr>
</thead>
</table>

**Select Participants**
- Age at
  - baseline
  - 0-17
  - 18-64
  - 65+
- Year of birth
- Gender
  - Male
  - Female
- Subcohorts
  - GWAS
  - UG1
  - DEEP
  - DAG3

**Select Assessments**

**View data**

- Demographics
- Diseases & symptoms
- Elderly
- Family health
- Lifestyle & environment
- Medical treatment
- Mental health
- Nutrition
- Physical state
- Reproduction & development
- Secondary & linked data
- Wellbeing & subjective

**Emotional affect (PANAS)**
- panas_active_adu_q_19
- panas_afraid_adu_q_20
- panas_alert_adu_q_12
- panas_ashamed_adu_q_13
- panas_attentive_adu_q_17
- panas_determined_adu_q_16
- panas_distressed_adu_q_02
- panas_enthusiastic_adu_q_09
- panas_excited_adu_q_03
- panas_guilty_adu_q_06
- panas_hostile_adu_q_08
- panas_inspired_adu_q_14
- panas_interesting_adu_q_01
- panas_irritable_adu_q_11
- panas_jittery_adu_q_18

20 variables found in subsection "Emotional affect (PANAS)".

Please sign in to select and order variables.

https://data-catalogue.lifelines.nl/
data release 2.0

data order in catalogue results in pseudonomized tables

- **automatically** generated via platform
- 1 table per element
  - includes list of participants + variants
  - if variable was assessed but empty -> ‘missing token’

- **additional table 1:** variable metadata
- **additional table 2:** global summary per participant:
  - DOB/DOD
  - intervals between elements (in months)
  - invited for elements yes/no
Mini International Neuropsychiatric Interview

The Mini International Neuropsychiatric Interview (MINI) was performed in all adult Lifelines participants during:

- 2A Visit 1 (as interview, n = ~149,000, 65+ participants only with MMSE score of 26 or more)
- 2A Visit 1 (as a digital questionnaire on location, n = ~119,000)
- 3A Visit 1 (as a digital questionnaire at home)

Note that MINI derivatives were calculated to assess whether participants meet the criteria for certain psychiatric diagnoses.

Background

The Mini International Neuropsychiatric Interview (MINI) is a brief, reliable and valid structured diagnostic interview for diagnosing psychiatric disorders, compatible with international diagnostic criteria such as the DSM-IV and ICD-10.

The MINI was performed in adult and elderly Lifelines participants as a face-to-face interview with a nurse practitioner on location during 1A Visit 1, and as a digital questionnaire on location during 2A Visit 1. In general assessment 3A, the MINI is implemented as a digital questionnaire to be filled in at home. The MINI should not be confused with the Mini Mental State Examination which is also used in Lifelines, but measures cognitive impairments. In addition to the MINI, which asks questions on altered mood during various time intervals, the additional assessment BIAQ was performed to assess lifetime mood-related episodes (LIDAS).

The MINI used in the Lifelines cohort (section: mental health) is restricted to the following therapeutic indications:

- major depressive episode
- dysthymia
- panic disorder
- agoraphobia
- social phobia
- generalized anxiety disorder

Versions of the MINI

In 1A Visit 1, three different versions of the MINI (face-to-face interview) were used.

ADHD (ADHQ)

ADHD is the name of an additional questionnaire on prevalence and co-morbidities of attention deficit hyperactivity disorder among Lifelines participants. Implementation of this questionnaire in Lifelines was initiated by the University of Amsterdam of the UMCG, as part of the EU Horizon-2020 CoCa (comorbid conditions of ADHD) project.

The collected ADHD data is currently under embargo. Please contact the Lifelines data managers at data@lifelines.nl for more information.

Protocol

ADHQ aimed to measure the rate of ADHD within the Lifelines cohort, as well as potential co-morbidities such as autism spectrum disorders, mood disorders, substance abuse, and sleeping problems. ADHQ contained separate but overlapping digital questionnaires that were tailored to three groups of Lifelines participants, plus a group of significant others:

1. Children 4-12 years old. Parents filled in the complete questionnaire. In this group, the questionnaire also includes a Developmental Coordination Disorder (DCD) questionnaire.
2. Adolescents 13-17 years old. Adolescents filled in their own questionnaire, except for items on ADHD and autism spectrum disorders which were filled in by their parents.
3. Adults of 18+ years old filled in their own questionnaires.
4. For adult participants, items on ADHD, cognition, autism spectrum disorders, and aggression were also included in a significant other such as a parent, partner, or similar.

Rationale for the use of parental or other informants: self-insight is limited for neurodevelopmental problems such as ADHD or autism spectrum, and a second informant is required to rate the presence of the symptoms in order to meet the existing publication standards regarding valid diagnostic information.

Timeline

- Batch 1: March - June 2017 (all age groups)
- Batch 2: Jan - Apr 2018 (additional adults and adult informants only)
- Batch 3: Jan - Apr 2019 (non-responders of adults and adult informants only)

Response

Children (4-12):
5,700 children (i.e., their parents) were invited for ADHQ.
QUESTIONS???
BBMRI Expert Meeting 26-11-2020

trynke de jong (t.r.de.jong@lifelines.nl)
This webinar will be recorded
Recordings can be made available for knowledge dissemination.

Questions or remarks?
Please raise your hand by clicking on the “raise your hand”-symbol or use the chat

Mute your microphone
Please mute your microphone to minimize background noise by clicking on the “mute”-symbol on the top of your screen.

What can we improve for the future?
We are happy to hear any additional questions or feedback for our next expert meeting. Please contact us by mail: robin.verjans@lygature.org
DataHub, towards a FAIR future

Visit us at
https://datahubmaastricht.nl
**Summary**

**DataHub is an institutional data broker**

made, managed and extended by our academic community

---

**Why, what**

- Science is in a phase of rapid transition due to the rise of digital technology, open science and the general data protection regulations (GDPR)
- DataHub is an institutional data broker. Made, managed and extended by our academic community

**Stick and carrot**

- A researcher’s heart lies in conducting research, the DataHub community is here to help
- Based on the condition “structure early, share later” the infrastructure can be best described as “a state-of-the-art GDPR compliant FAIR data station”

**Roadmap and next steps**

- Together with our partners we continuously improve our services in order to provide added value to researchers who want to do more with their data and data of colleagues
- Visit our portal, get in touch with us and, if needed, customize on-site by working with us on your use case
Science is in a phase of rapid transition due to the rise of digital technology and new forms of collaboration

- Open Science\(^1\): As a matter of principle, it is important that society as a whole should benefit from publicly funded research
- Go FAIR\(^2\): a bottom-up international approach for the practical implementation of the European Open Science Cloud (EOSC) as part of a global Internet of FAIR Data & Services

General Data Protection Regulation\(^3\) (GDPR): specify how data should be used and protected, and harmonizes data protection laws across the EU

- Accountability (demonstrate compliance)
- No legal basis (e.g. consent) = No data processing
- Privacy by Design and by Default
- Records of all data processing activities
- Data processing agreements
- Data Protection Officer (DPO)
- Far-reaching data subject rights

\(^1\) [https://www.openscience.nl/en](https://www.openscience.nl/en)
\(^2\) [https://www.go-fair.org/](https://www.go-fair.org/)
\(^3\) [https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/avg-nieuwe-europese-privacywetgeving/](https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/avg-nieuwe-europese-privacywetgeving/)
Mission
Linking community and technology to enable FAIR data

Small multidisciplinary core team
Research Software Engineers supplemented with embedded data stewards and possess the relevant domain knowledge

Characteristics
- Hospital and university!
- Data management (not data acquisition or science)!
- FAIR-inspired from start
- Structure early, share later
- Constantly improving self-service
- Use case based developments
- Parttime in kind on-site collaboration!
- Connected with the embedded data stewards
- Invest mostly in people versus software
- Open-source where possible
- Generic infrastructure, modular extendable
- Insourcing and outsourcing proven technology
Anatomy of a researcher

A researcher’s heart lies in conducting research, the DataHub community is here to help

- Funders require researchers to write a Data Management Plan (DMP) when they receive funding
  - (Pre-filled) DMP templates for researchers and a local DMP\(^1\) collaboration tool (based on DMP-online)
  - Researchers don’t need to find out about policy and legislation all by themselves
  - Organisation is more in control about e.g. the risk for data leaks

- Data governance
  - Compliant with (patient) privacy policies and de-identification guidelines
  - Translate the GDPR into (automated) concrete instructions

- We are here to help
  - Easy access to one stop shop\(^2\) concerning Research Data Management
  - Easy to use ‘state of the art’ infrastructure, made for collaboration and modular extendable
  - GDPR compliant FAIR data station

\(^1\) https://dmp.datahubmaastricht.nl/
\(^2\) https://datahubmaastricht.nl
Carrot

Based on the condition “structure early, share later” the infrastructure can be best described as “a state-of-the-art GDPR compliant FAIR data station[3]”

- Ingest & safely store research data[1]
  - Findable: Explore linked (meta)data
  - Accessible: Project data structuring, data lake, hierarchical organization in projects and datasets
  - Interoperable: (Meta)data structuring + ontology enrichment
  - Re-usable: Share (meta)data (after embargo period) and communicate provenance

- Pseudonymization services
  - Master Person Index is a GDPR cornerstone (algorithm to connect human data)
  - Trusted Second Party
  - (Double) pseudonymization
  - Automated pipelines

- DataHub catalogue
  - Based on linked (meta)data
  - Internal (protected) data linked with external (open) data[2]
  - Federated search capabilities improving the discovery of relevant data

Complex queries like:
"What are the companies producing antimetabolite drugs used in phase 2 clinical trails where patients can still enroll to treat lung cancer and where EGFR is mentioned" resolved within minutes instead of ...

[1] https://rdm.datahubmaastricht.nl/
Challenges
"Not all data is equal."

Types of data:
- Structured
- Unstructured
- Personal
- Non-Personal
- Meta-data

8 types of data
8 TYPES OF DATA

OUTLOGIES

DEPERSONALIZATION

HARD

STRUCTURED PERSONAL META-DATA

STRUCTURED PERSONAL DATA

STRUCTURED NON-PERSONAL META-DATA

STRUCTURED NON-PERSONAL DATA

UNSTRUCTURED PERSONAL META-DATA

UNSTRUCTURED PERSONAL DATA

UNSTRUCTURED NON-PERSONAL META-DATA

UNSTRUCTURED NON-PERSONAL DATA

WHERE TO START?

VS
THE BIG PROMISE OF DATA

PERSONAL MEDICINE

INNOVATION

REPOSITORIES

STRUCTURING

ONTOLOGIES

INTERFACES

OPEN DATA FORMATS

COLLABORATION TOOLS

DEPERSONALIZATION

STORAGE

META DATA

RE-USABLE

INTEROPERABLE

ACCESSIBLE

FINDABLE

WHAT DOES THE JOURNEY LOOK LIKE?

WHAT DOES IT TAKE TO GET TO NEXT LEVEL?

WHERE TO START?

EASY DATA FIRST

SUPPORT ALL TYPES OF DATA

WHERE TO START?
Visit us at https://datahubmaastricht.nl

Maastricht UMC+
DataHub

Pascal Suppers
Managing Director

DataHub Maastricht
P. Debyelaan 15
6229 HX Maastricht
The Netherlands

L. van Kleeftoren
2nd floor (route 11)
The Netherlands

T +31 6 27 07 16 54
E p.suppers@maastrichtuniversity.nl
THANK YOU FOR ATTENDING OUR EXPERT MEETING SERIES!

FEEDBACK?! → robin.verjans@lygature.org
• Did the expert meetings live up to your expectations?
• What did you gain?
• What did you miss?

2020
Examples of data sharing procedures, tools & challenges

2021
In-depth discussion themes:
1. Data Findability
2. Data Request Management
3. Data Linkage
4. Legal Framework
5. ...What do you prefer to discuss?...