



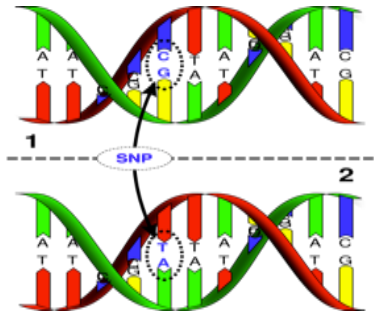
Personalised medicine: *Hoe om te gaan met de data?*

Salome Scholtens
April 2018

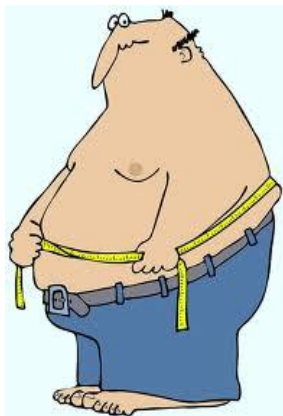
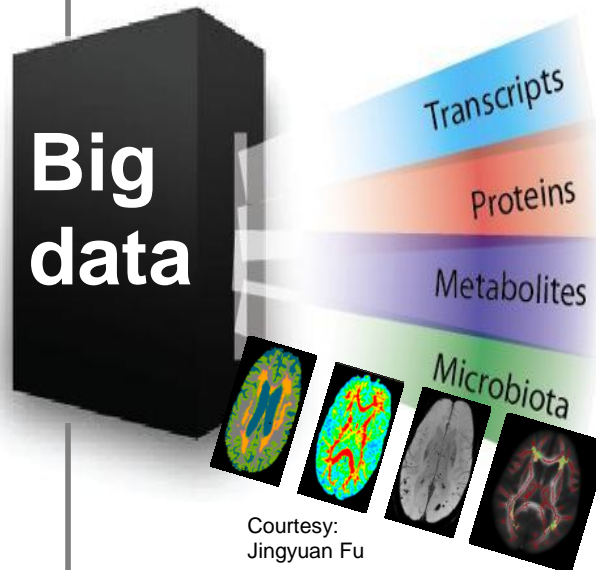
Disclosure belangen spreker

(potentiële) belangenverstrengeling	Geen
Voor bijeenkomst mogelijk relevante relaties met bedrijven	
<ul style="list-style-type: none">• Sponsoring of onderzoeksgeld• Honorarium of andere (financiële) vergoeding• Aandeelhouder• Andere relatie, namelijk ...	<ul style="list-style-type: none">••••

Personalised Medicine *Data*



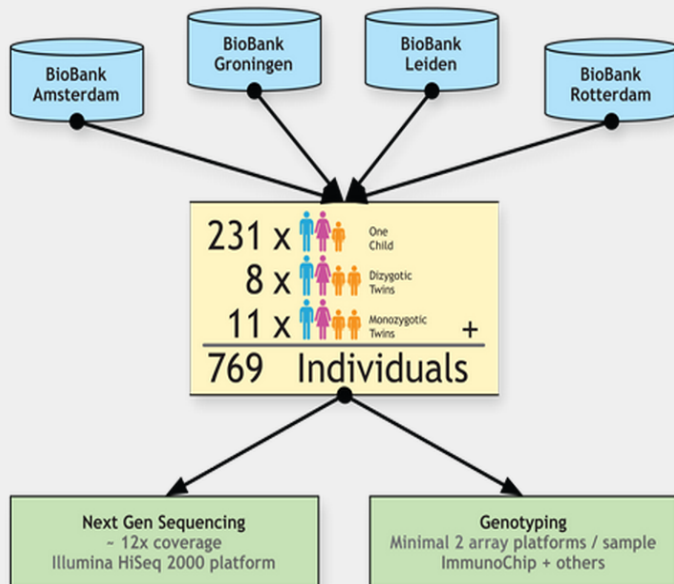
Welke DNA varianten (genen) beïnvloeden ontstaan en verloop van ziekte ?



Waarin verschillen mensen in ontstaan en verloop van ziekte. Wat is de rol van omgeving?

Ultra-sharp genetic group portrait of the Dutch

What genetic variation is to be found in the Dutch indigenous population? Detailed knowledge about this is not only interesting in itself, it also helps to extract useful biomedical information from Dutch biobanks. The Dutch biobank collaboration BBMRI-NL has initiated the extensive Rainbow Project "Genome of the Netherlands" (GoNL) because it offers unique opportunities for science and for the development of new treatments and diagnostic techniques. A close-up look at the DNA of 750 Dutch people-250 trio's of two parents and an adult child-plus a global genetic profile of large numbers of Dutch will disclose a wealth of new information, new insights, and possible applications.



Dutch reference set

Boomsma et al, 2013, EJHG
 Genome of the Netherlands consortium, 2014, Nat Gen

No. samples	769
DNA source	Blood
Coverage	14.5x
Insert	500bp
Readlength	90bp
Data generation	BGI/Illumina
Population	Dutch only, trios, twins
QC	GWAS arrays, Trios
Phenotype info	Multiple

Enorme datasets

- 45TB ruwe data (fq.gz)
- 450TB intermediate data (bam)
- 90TB result (bam + vcf)

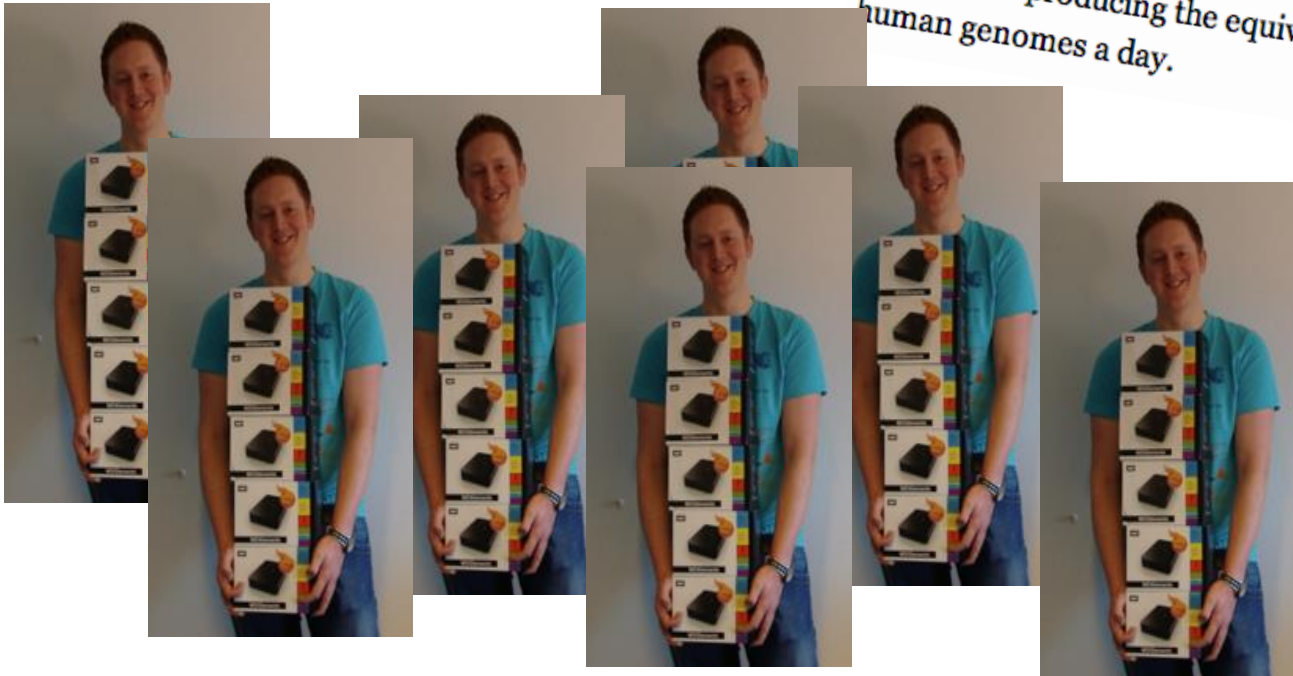
The New York Times

NEW YORK, FRIDAY, NOVEMBER

DNA Sequencing Caught in Deluge of Data
BGI, based in China, is the world's largest
genomics research institute, with 167 DNA
sequencers producing the equivalent of 2,000
human genomes a day.

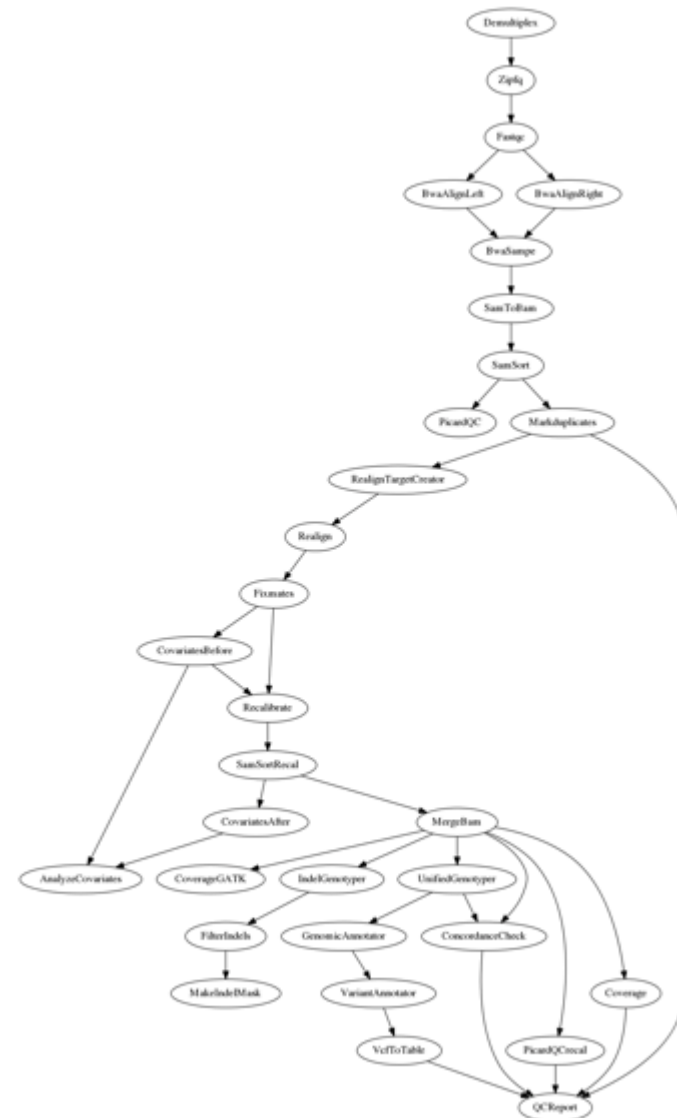


Freerk van Dijk, of the Netherlands
BGI
Genome Project

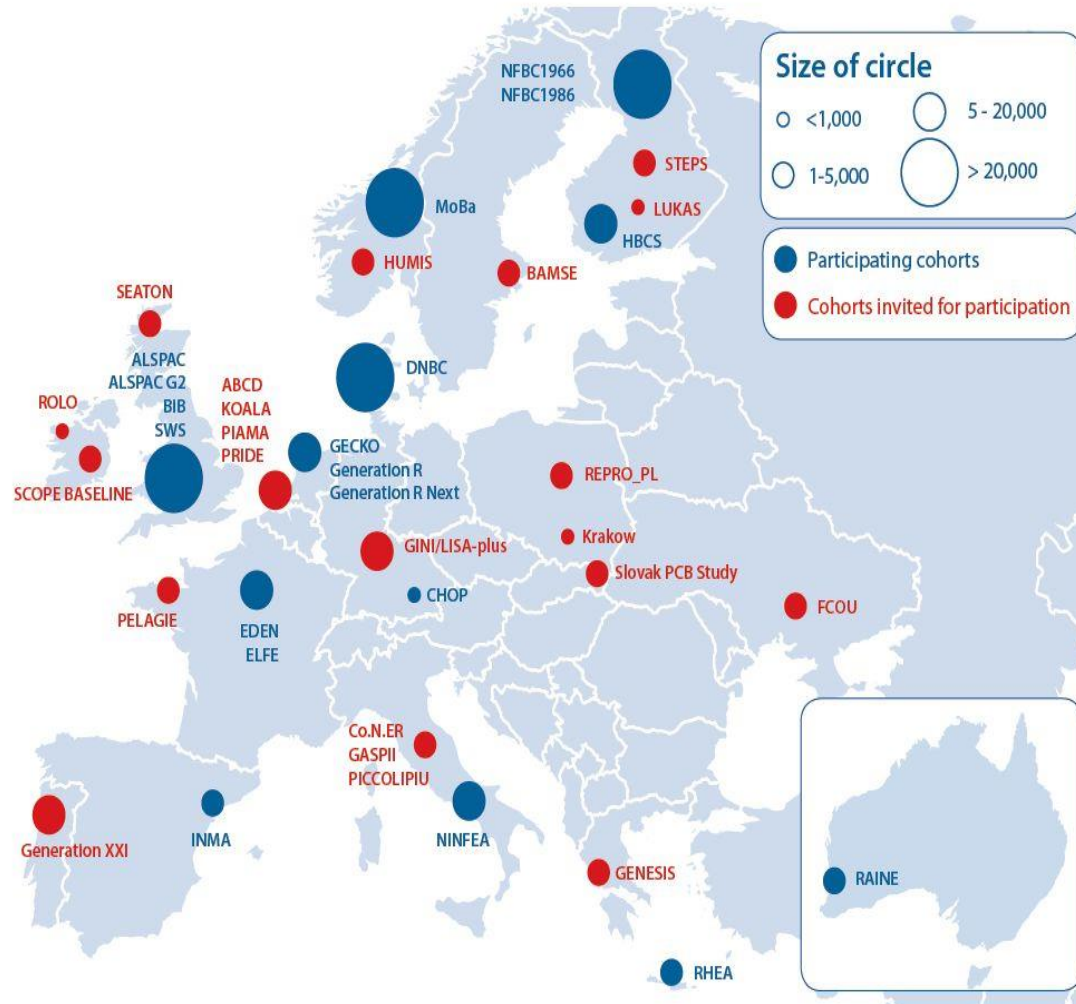


>30 analyse stappen nodig

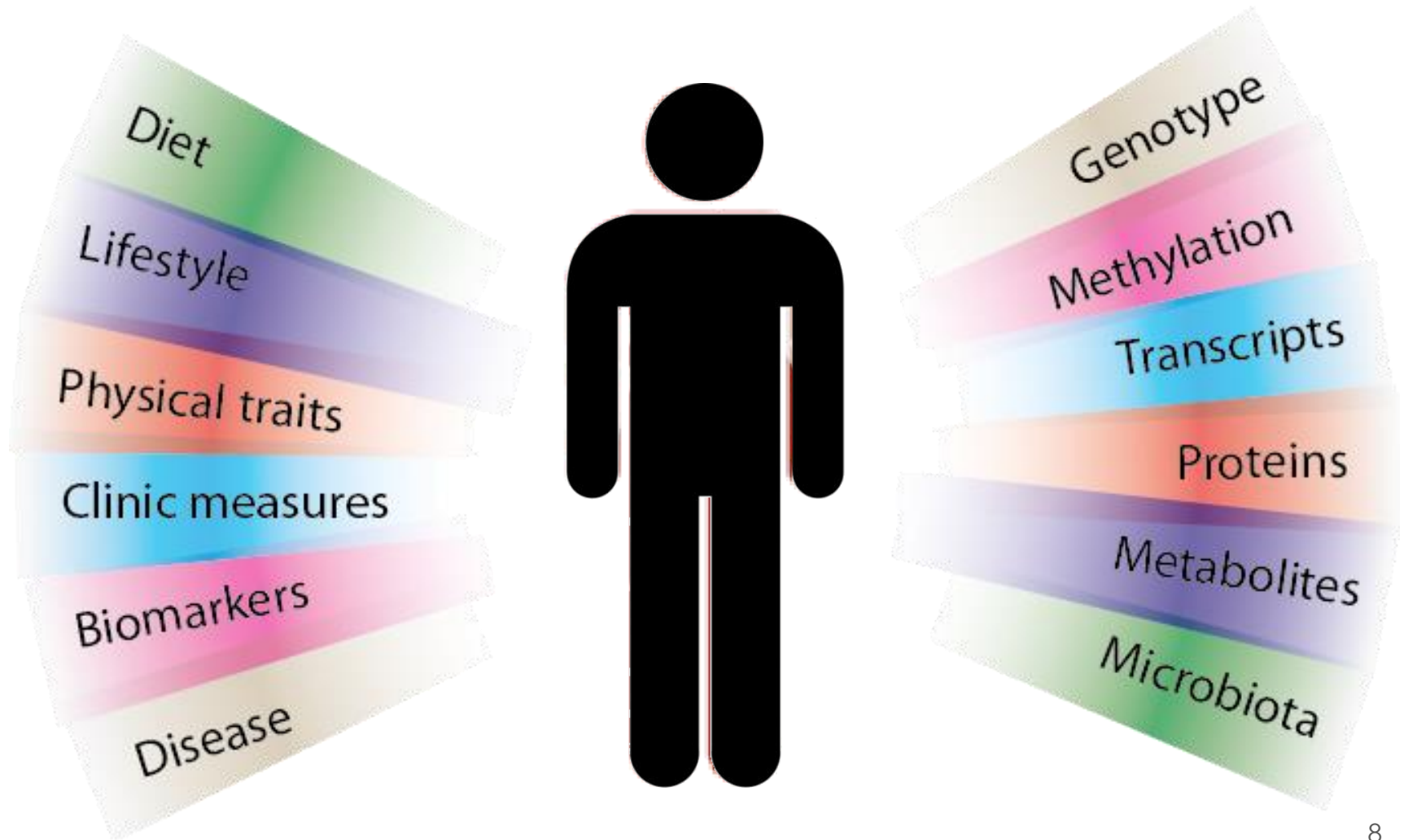
1	name	protocol_name	PreviousSteps_name
2	Demultiplex	Demultiplex	
3	Zipfq	Zipfq	Demultiplex
4	Fastqc	Fastqc	Zipfq
5	BwaAlignLeft	BwaAlignLeft	Fastqc
6	BwaAlignRight	BwaAlignRight	Fastqc
7	BwaSampe	BwaSampe	BwaAlignLeft, BwaAlignRight
8	SamToBam	SamToBam	BwaSampe
9	SamSort	SamSort	SamToBam
10	PicardQC	PicardQC	SamSort
11	Markduplicates	Markduplicates	SamSort
12	RealignTargetCreator	RealignTargetCreator	Markduplicates
13	Realign	Realign	RealignTargetCreator
14	Fixmates	Fixmates	Realign
15	CovariatesBefore	CovariatesBefore	Fixmates
16	Recalibrate	Recalibrate	Fixmates, CovariatesBefore
17	SamSortRecal	SamSortRecal	Recalibrate
18	CovariatesAfter	CovariatesAfter	SamSortRecal
19	AnalyzeCovariates	AnalyzeCovariates	CovariatesBefore, CovariatesAfter
20	PicardQCrecal	PicardQCrecal	SamSortRecal
21	Coverage	Coverage	SamSortRecal
22	IndelGenotyper	IndelGenotyper	SamSortRecal
23	FilterIndels	FilterIndels	IndelGenotyper
24	UnifiedGenotyper	UnifiedGenotyper	SamSortRecal
25	MakeIndelMask	MakeIndelMask	FilterIndels
26	VariantAnnotator	VariantAnnotator	SamSortRecal, UnifiedGenotyper
27	GenomicAnnotator	GenomicAnnotator	VariantAnnotator
28	VariantFiltration	VariantFiltration	GenomicAnnotator, MakeIndelMask
29	VcfToTable	VcfToTable	VariantFiltration
30	ConcordanceCheck	ConcordanceCheck	SamSortRecal, UnifiedGenotyper
31	QCReport	QCReport	PicardQC, ConcordanceCheck



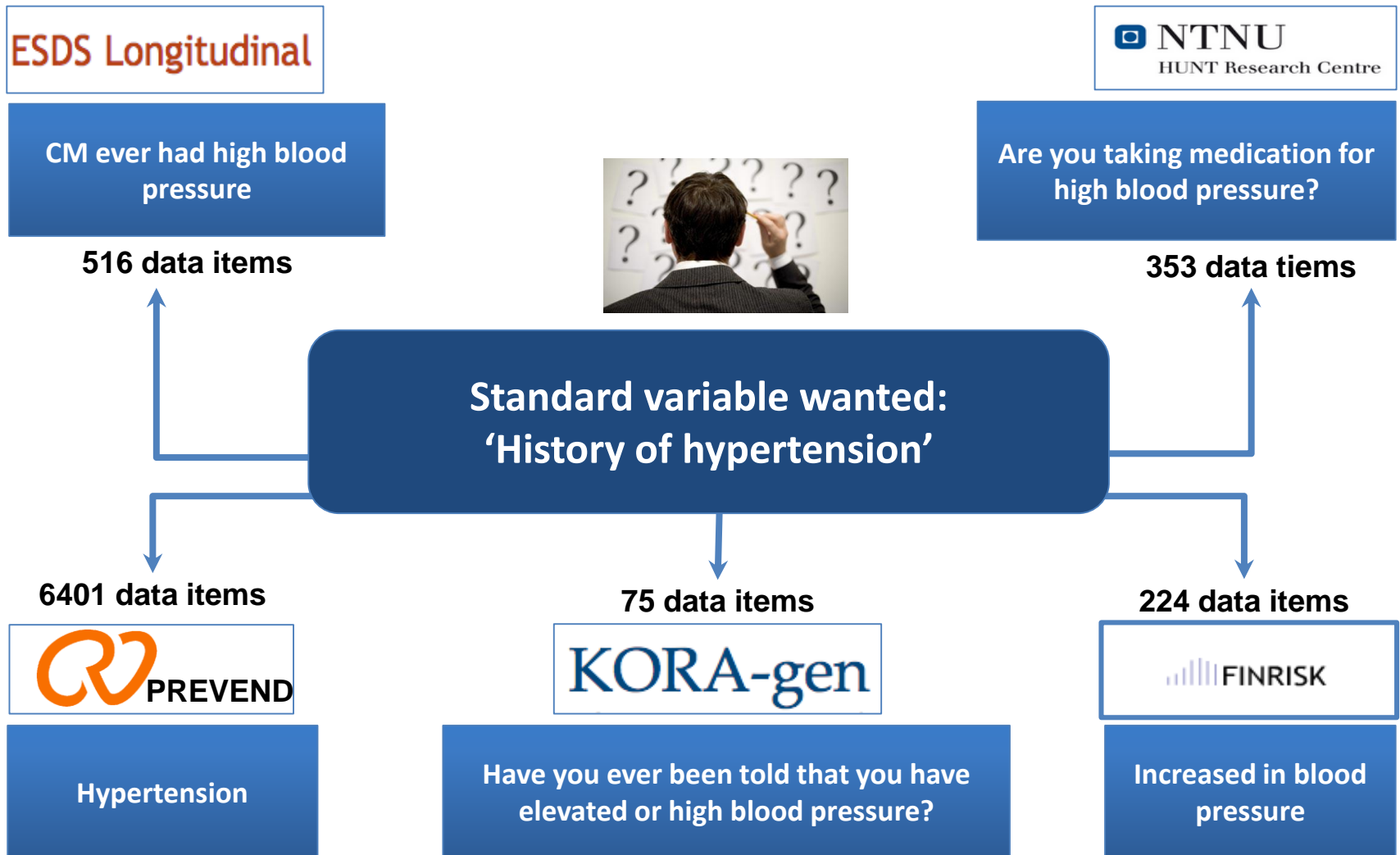
Data niet op 1 plek



Multi-layer data

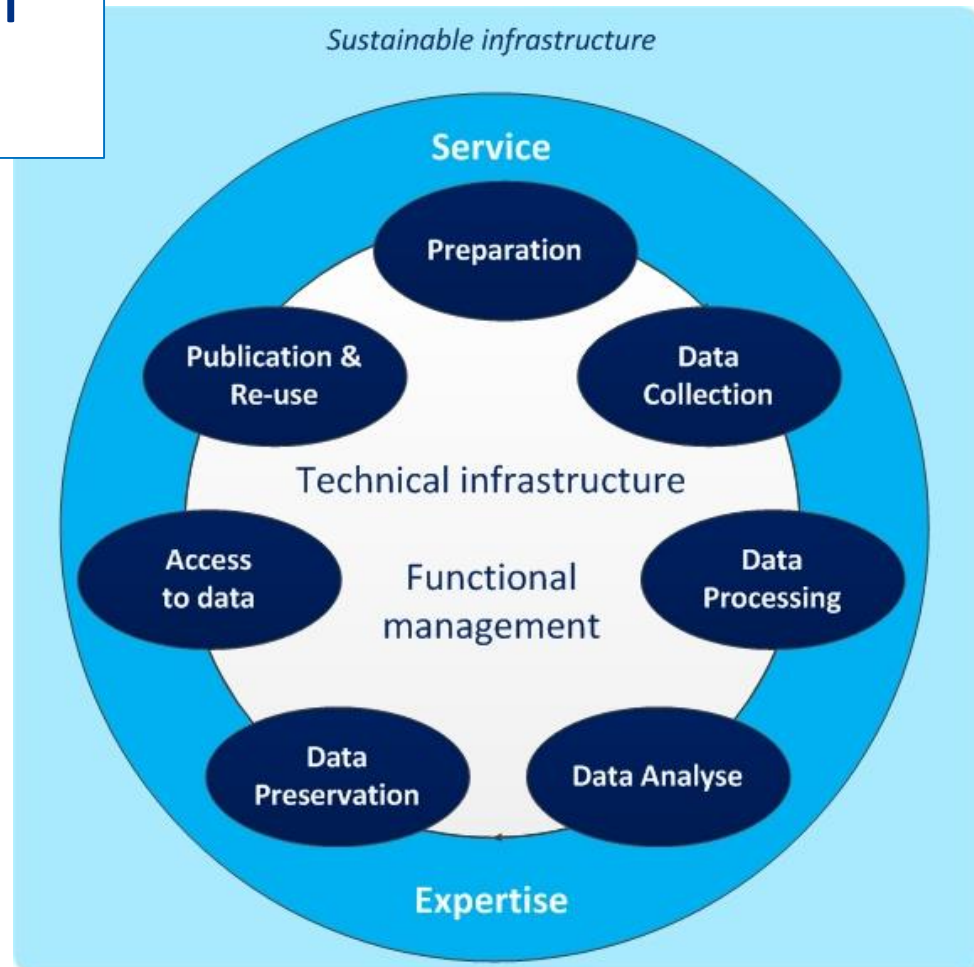


Data harmonisation

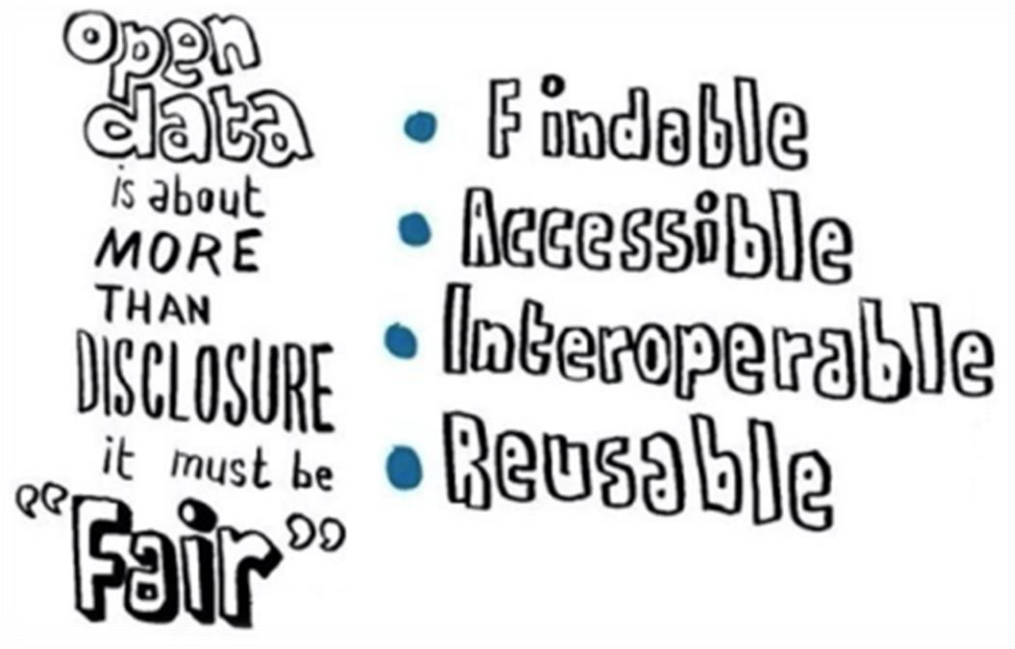


Data Life cycle

Data is constant in
beweging....



FAIR data



- **Findable** : Makkelijk te vinden en te begrijpen
- **Accessible**: Goed beschreven condities voor data toegang
- **Interoperable**: Zodat data kan worden gecombineerd
- **Re-Usable**: Voor toekomstig onderzoek en replicatie

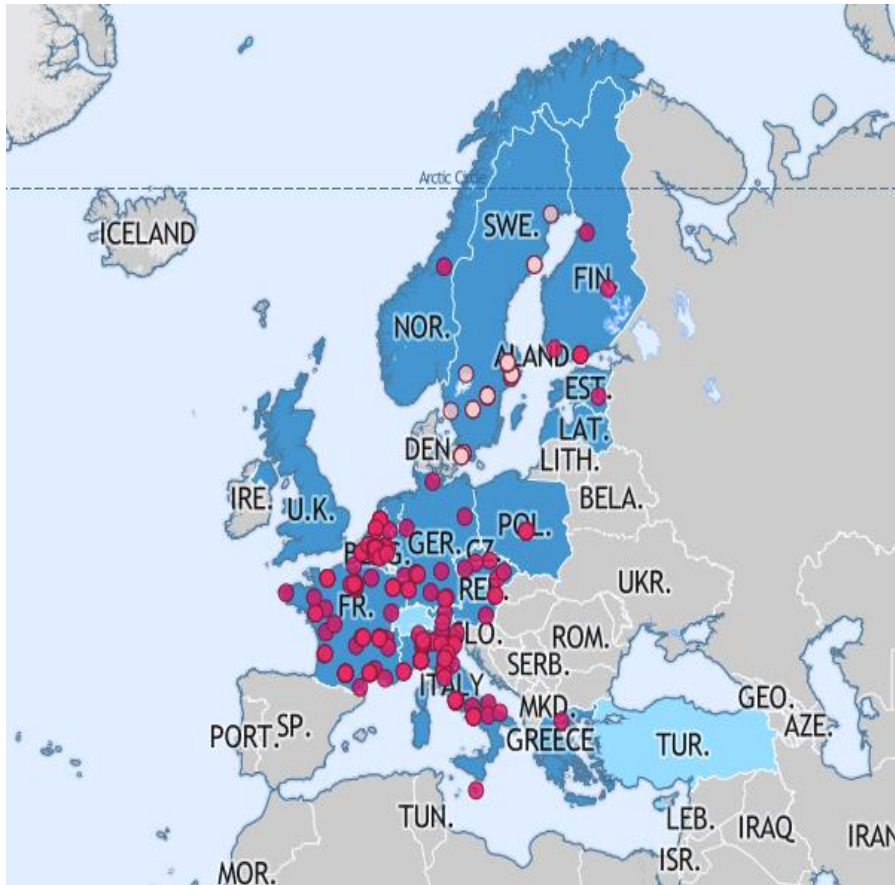


Voor effectief onderzoek:

- Maak een plan!
 - ZonMW data management plan
- Zorg voor passende tools en infrastructuur
 - Veilig data warehouse systeem
 - Platform voor delen/combineren van data
 - Vinden van data: catalogi
- Support
 - Data steward onderdeel van onderzoeksteam
 - Bioinformatica / data science support
- Training
 - Onderzoekers
 - Support staff



Findable data: catalogi



<http://catalogue.bbmri.nl>

239 collections + datasets

<http://catalogue.lifelines.nl>

~1000 data items

<http://www.palgaopenbaredatabank.nl/>

35.736.175 reports (\geq samples)

<http://directory.bbmri-eric.eu>

~1400 collections, 189 dutch

<https://samples.rd-connect.eu/>

~ 7,5k Rare Disease Samples

www.groningendatacatalogus.nl

Local collection, connected with national

Nationale Biobank catalogus



B B M R I • N L

Catalogue

Feedback

Admin

Sign in

Sample Collections

Sample Collections

DNA RNA



Data item filters



Data item selection

Select all

Deselect all

- Name
- Acronym
- Collection type
- Diseases studied
- Data categories
- Material collected
- Experimental data
- Gender
- Age low
- Age high
- Age unit
- Number of donors
- Description
- Publications
- Organisational structure
- Sample access policy
- Data access policy

Data

Aggregates

Name	Acronym	Collection type	Material collected
N/A		Disease specific	cDNA / mRNA
U-TWIN		Disease specific	DNA,cDNA / mRNA,Whole Blood
MINDACT		Disease specific	DNA,cDNA / mRNA,Other
N/A		Disease specific	DNA,Tissue, paraffin preserved,cDNA / mF
Cohort on Diabetes and Atherosclerosis	CODAM	Disease specific	DNA,cDNA / mRNA,Plasma,Serum
Severely Obese Individuals Maastricht		Disease specific	DNA,cDNA / mRNA
Netherlands Twin Register	NTR	Population-based	DNA,Urine,cDNA / mRNA,Plasma,Serum
Generation R		Population-based	DNA,Urine,cDNA / mRNA,Plasma,Serum
Erasmus Rucphen Family study (ERF) and patient cohorts: family-based		Population-based	DNA,Urine,cDNA / mRNA,Plasma,Serum
NESDA		Disease specific	DNA,Saliva,cDNA / mRNA,Plasma,Serum
GARP + RAAK study		Disease specific	DNA,cDNA / mRNA,Other,Whole Blood,Pl
Parel Leukemie, Myeloom, Lymfeklierkanker	Parel LML	Disease specific	DNA,cDNA / mRNA,Other,Whole Blood,Se
ERGO, Rotterdam study: population-based; outbred		Population-based	DNA,Saliva,Urine,cDNA / mRNA,Feeces,Pl
NELSON	NELSON	Disease specific	DNA,cDNA / mRNA,Plasma,Serum

14 items found

Download

Aanvraag procedure

Submit Study Request

Project title *

Request form *

Variable	Description	Remove
Hip	Hip circumference in cm	✕
Length	Length in cm	✕
Position	Standing or sitting	✕
Span	Span in cm	✕
Waist	Waist circumference in cm	✕
Waist sitting	Waist circumference in cm (sitting)	✕
Weight	Weight in kg	✕
Weight reported at visit	Weight according to participant in kg	✕

HPC infrastructuur



B B M R I • N L



Voor effectief onderzoek in Personalised Medicine:

- Zorg dat je weet hoe je goed met data omgaat
- Opleiden van onderzoekers en ondersteuner
- Onderdeel van opleiding aio's
- Facultatieve verdiepende modules
- In samenwerking met ZonMW:
 - Hands-on FAIR data training
 - Train-de-trainer





ZonMw



BBMRI-ERIC



health

RI research
infrastructure

DTL



DUTCH TECHCENTRE FOR LIFE SCIENCES



trait

a ctmm project



NFU

NEDERLANDSE FEDERATIE VAN
UNIVERSITAIR MEDISCHE CENTRA

Vragen?

s.scholten@umcg.nl