

**Inter  
Connect**



*Global data for diabetes and obesity research*

# **Symposium: Global data for diabetes and obesity research**

***Host: Nick Wareham***

***InterConnect Co-ordinator & Director, MRC Epidemiology Unit, University of Cambridge, UK***

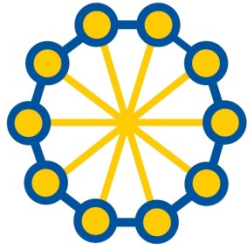
***Venue: EASD Stockholm, 14<sup>th</sup> September 2015***

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

# Programme

---

<b>14:30</b>	<b>Welcome</b>	Nick Wareham
<b>14:35</b>	<b>Scientific opportunity, challenge and vision</b>	
14:35	Understanding differences in risk of diabetes and obesity between populations	Nick Wareham
14:50	Challenges of data sharing models	Nita Forouhi
15:10	InterConnect: Vision of a changed paradigm	Nick Wareham
<b>15:30</b>	<b>Delivering the InterConnect vision</b>	
15:30	Data discovery: The registry	Matthias Schulze
15:45	Bringing the analysis to the data: Proof of concept	Tom Bishop
16:00	Developing the vision via exemplar research questions	Ken Ong
<b>16:15</b>	<b>Open discussion and involvement</b>	Nick Wareham



**Inter**  
**Connect**



*Global data for diabetes and obesity research*

# Understanding differences in risk of diabetes and obesity between populations

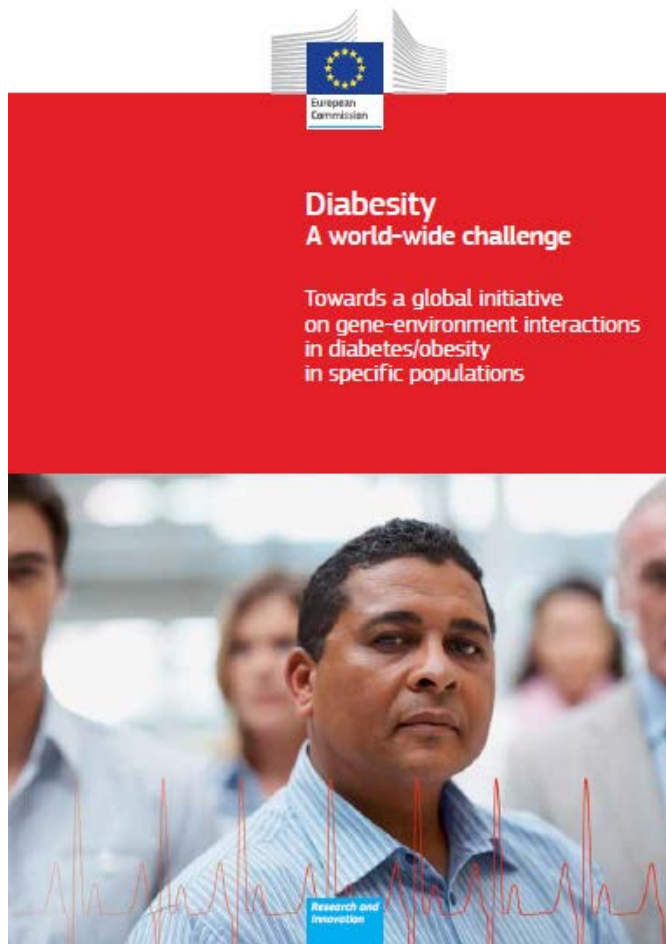
*Nick Wareham*

*InterConnect Co-ordinator & Director, MRC Epidemiology Unit,  
University of Cambridge, UK*

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

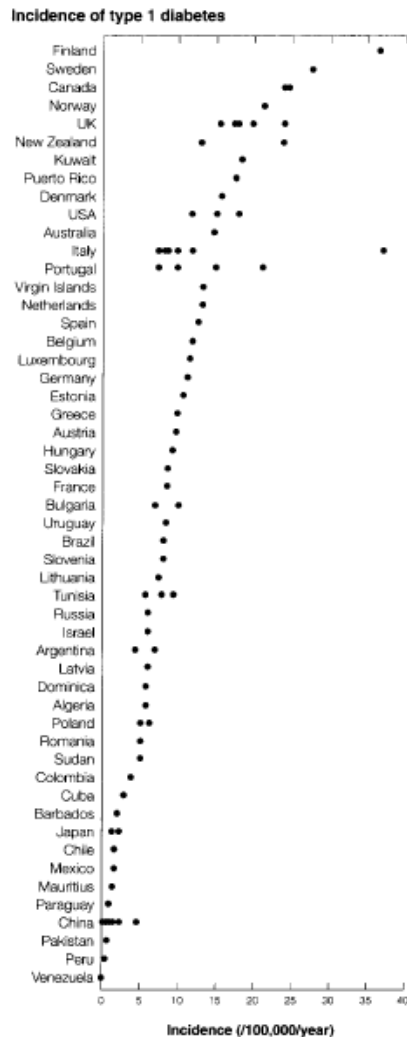
# EU “diabesity” conference 2012

---



- Research into individual and societal approaches to the prevention of obesity, diabetes and related metabolic disorders
- Health systems interventions to better treat diabetes
- Research into understanding differences in individual and population risk

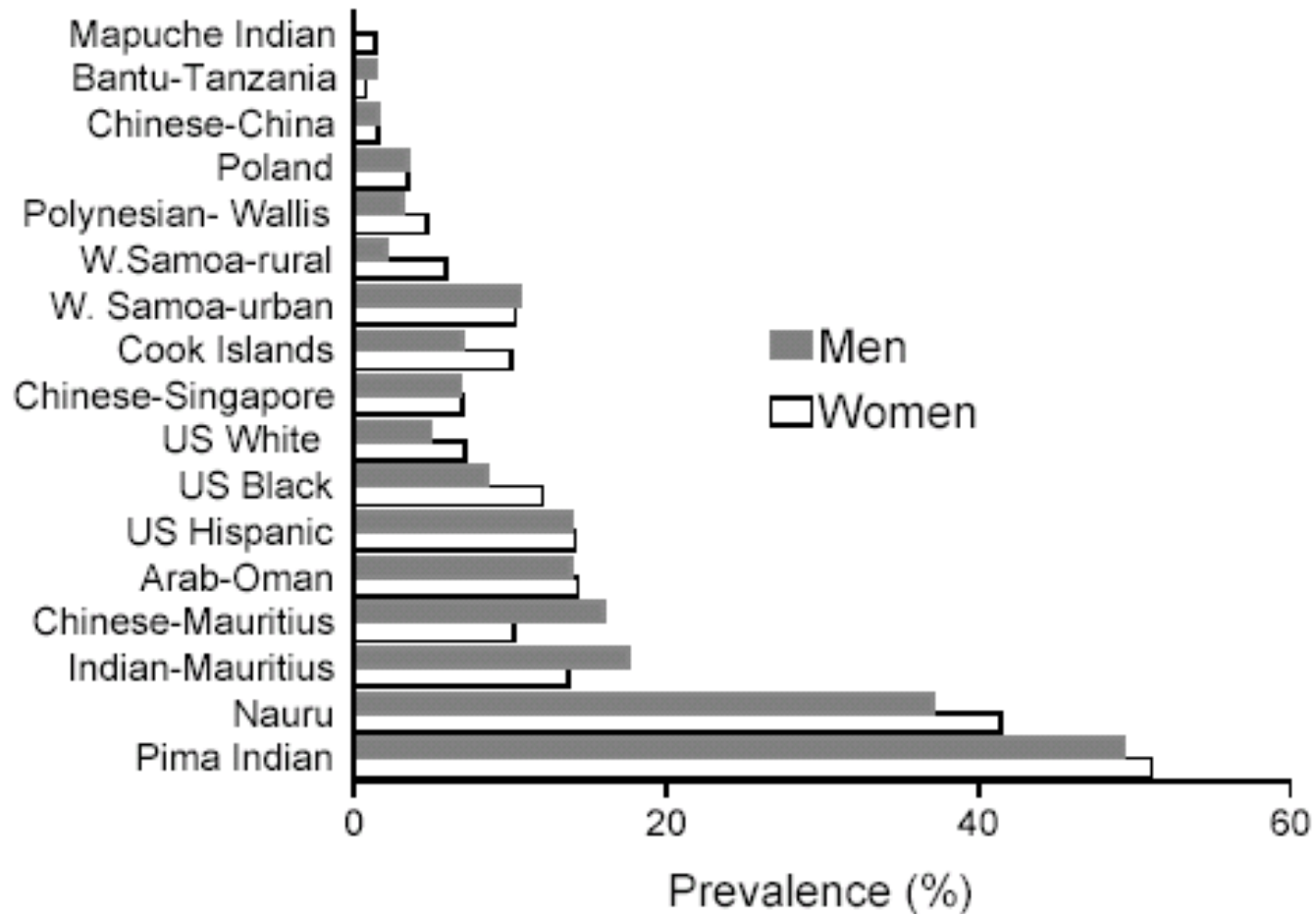
# Between-population differences in incidence of type 1 diabetes



- High incidence in Finland, Sardinia and other populations
- On-going cohort studies in specific populations investigating interplay between genetic susceptibility and environmental triggers

# Between-population differences in type 2 diabetes prevalence

---



# Possible explanations for between-population differences in prevalence

---

*THE AMERICAN JOURNAL*  
*of*  
HUMAN  
GENETICS

## Diabetes Mellitus: A “Thrifty” Genotype Rendered Detrimental by “Progress”?

JAMES V. NEEL  
*Department of Human Genetics,  
University of Michigan Medical School,  
Ann Arbor, Mich.*



**Source:** Neel, Am J Human Genetics 1962

# Possible explanations for between-population differences in prevalence

## Review

### Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis\*

C. N. Hales<sup>1</sup> and D. J. P. Barker<sup>2</sup>

<sup>1</sup> Department of Clinical Biochemistry, Addenbrooke's Hospital, Cambridge, and

<sup>2</sup> MRC Environmental Epidemiology Unit, University of Southampton, Southampton General Hospital, UK



Weight at Birth.	Weight 1st Year	Food.	No. of Visits.	Condition, and Remarks of Health Visitor.			
				W	V	P	T
8 7/8 lbs	24 1/2 lbs	B.	11	Y	-	-	4
Healthy & well developed.				Buckland School. Card to S.			
7 lbs	15 1/4 lbs	B	12	h.	Y.	Y.	8
Moved to Bury Green L. Southampton.				Had measles, pneumonia & c.			
8	20	Bol.	11	Y.	Y.	?	4
I.B. above in neck pinned. Int. foranella still open 23 yrs. Abdomen very large & p.							
8 1/2	22	B.B.	9	Y	Y	Y	10
Healthy & normal.				Buckland School. Card.			

Source: Hales and Barker, Diabetologia 1992



# Phase 2: Studying explanations for differences in risk between individuals within-populations



- EPIC-InterAct  
Nested case-cohort study within EPIC Europe
- Large  
455,680 individuals at baseline
- Long follow-up
  - 4 million person years
  - 12,403 incident cases of T2DM
- Stored blood
- Data on diet/physical activity
- Exposure heterogeneity

**Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study**

The InterAct Consortium



Research groups in 8 countries; 26 centres

**Source:** Langenberg C et al, Diabetologia 2011

# InterAct findings – foods associated with increased risk of T2DM



Diabetologia (2013) 56:47–59  
DOI 10.1007/s00125-012-2718-7

ARTICLE

## Association between dietary meat consumption and incident type 2 diabetes: the EPIC-InterAct study

The InterAct Consortium



Diabetologia (2013) 56:1520–1530  
DOI 10.1007/s00125-013-2899-8

ARTICLE

## Consumption of sweet beverages and type 2 diabetes incidence in European adults: results from EPIC-InterAct

The InterAct consortium

# InterAct findings – foods associated with reduced risk of T2DM



The amount and type of dairy product intake and incident type 2 diabetes: results from the EPIC-InterAct Study<sup>1-3</sup>

*Am J Clin Nutr* 2012



The prospective association between total and type of fish intake and type 2 diabetes in 8 European countries: EPIC-InterAct Study<sup>1-3</sup>

*Am J Clin Nutr* 2012

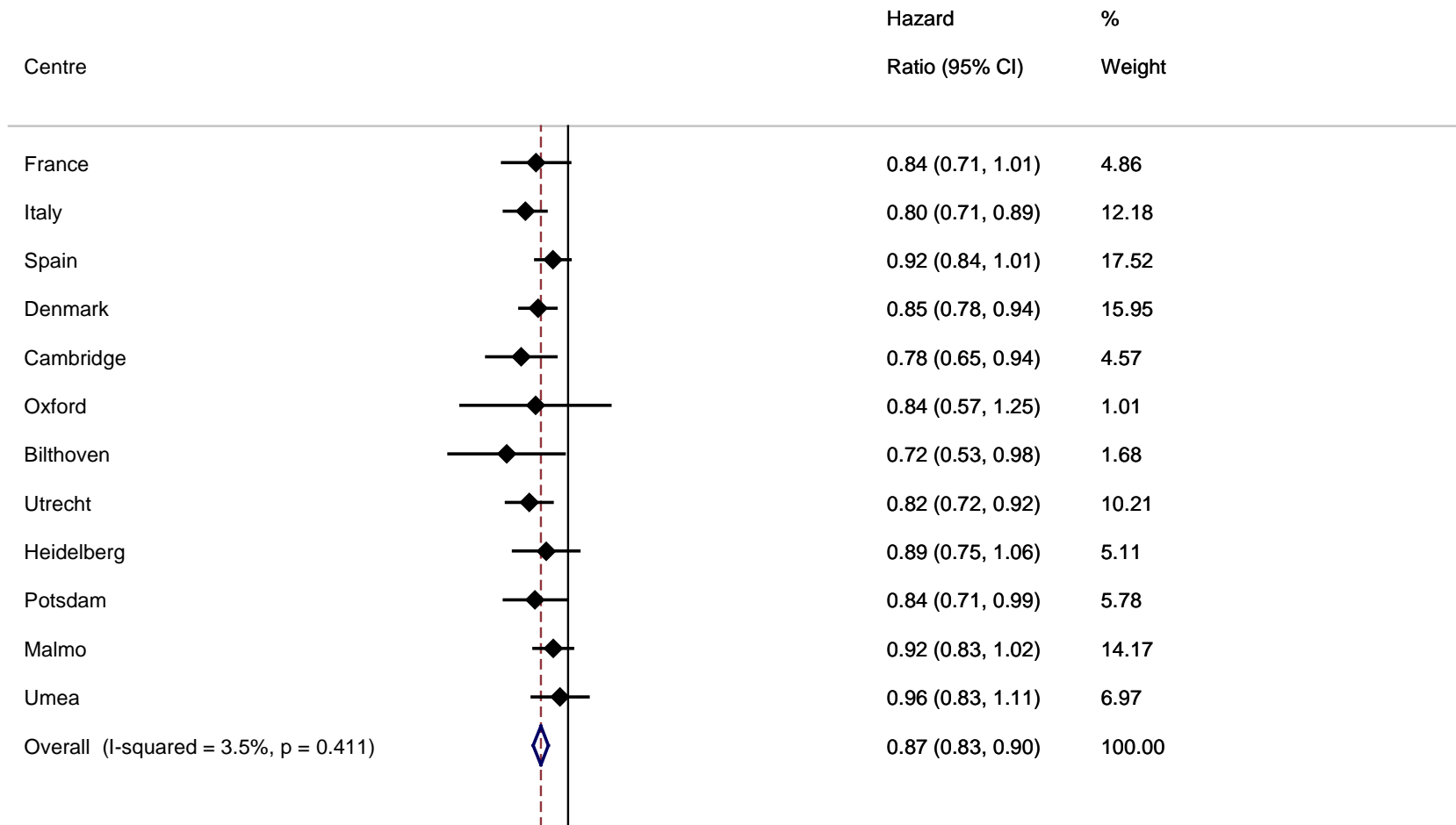
## SYSTEMATIC REVIEW

Fruit and vegetable intake and type 2 diabetes: EPIC-InterAct prospective study and meta-analysis

European Journal of Clinical Nutrition (2012)



# InterAct findings - Physical activity and risk of T2DM



Source: Ekelund et al, Diabetologia 2012

# InterAct findings: Main genetic effect of known variants

---



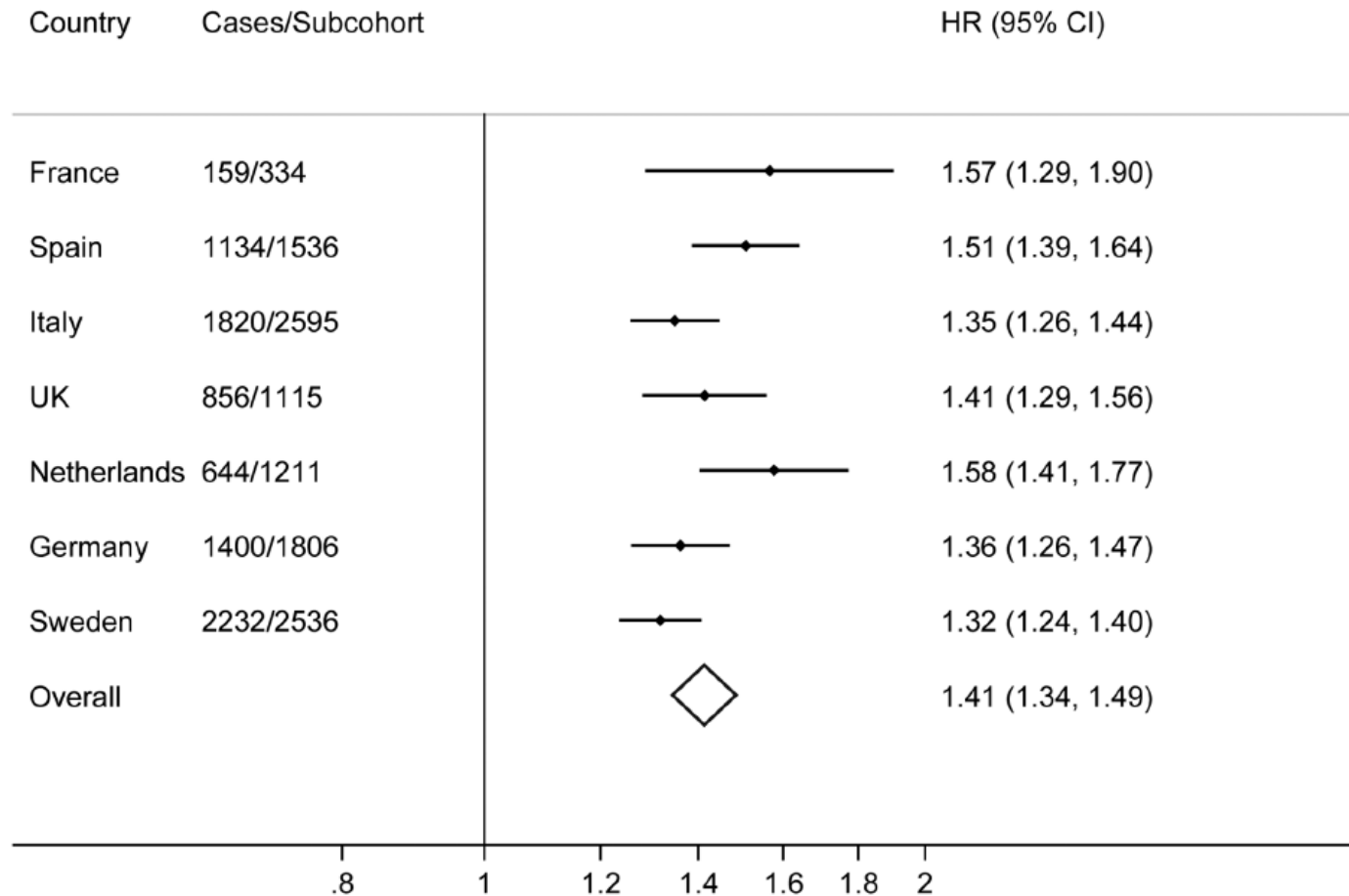
49 variants previously demonstrated to be associated with T2DM

Genetic risk score strongly associated with incident T2DM  
– HR per allele 1.08 (1.07-1.10)  $p = 10^{-41}$

Per SD of GRS HR = 1.41 (1.34-1.49)  $p = 10^{-41}$

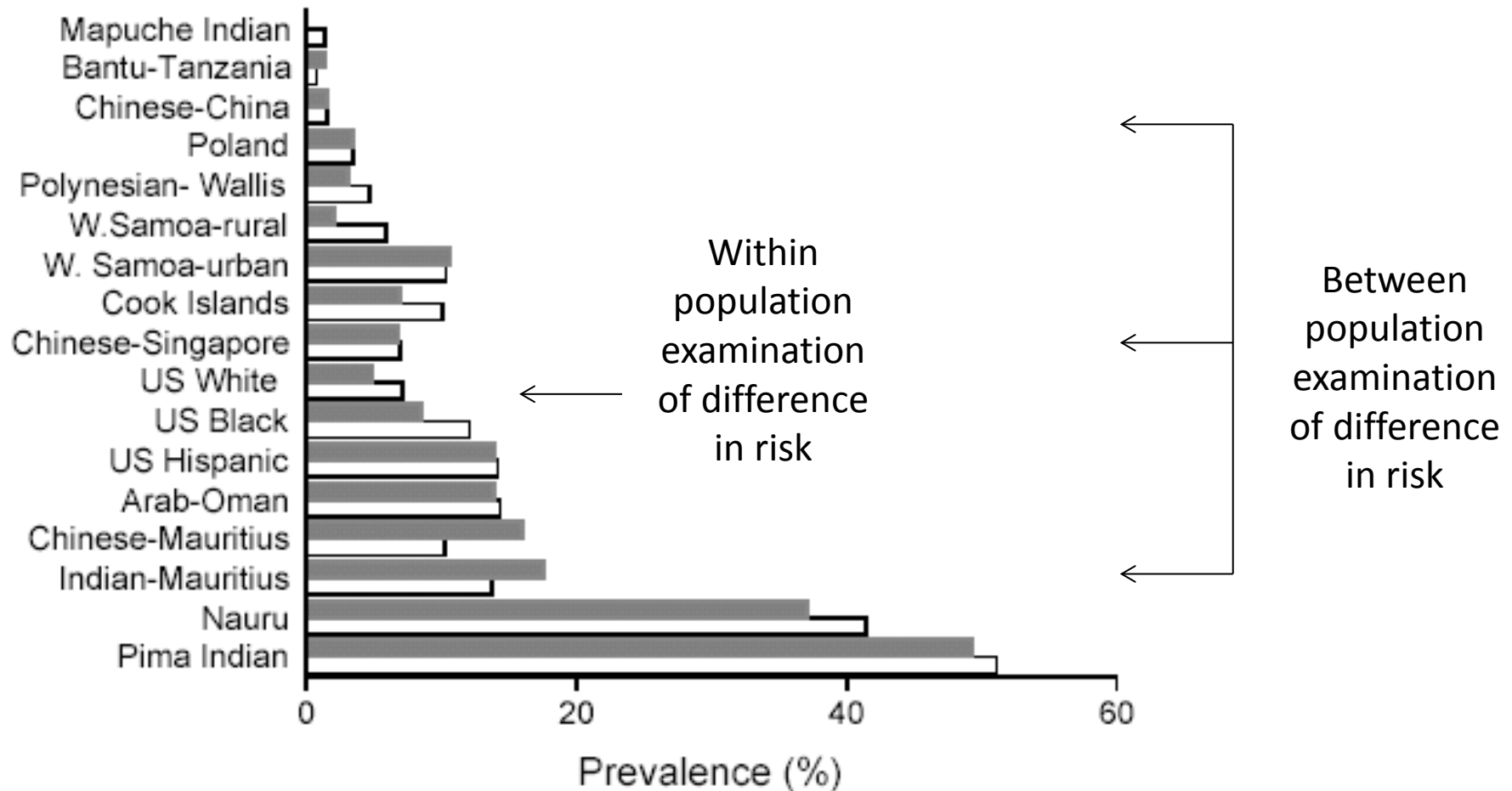
No evidence of interaction for individual gene variants with age, sex, family history, BMI or physical activity

# InterAct findings: Main genetic effect by country



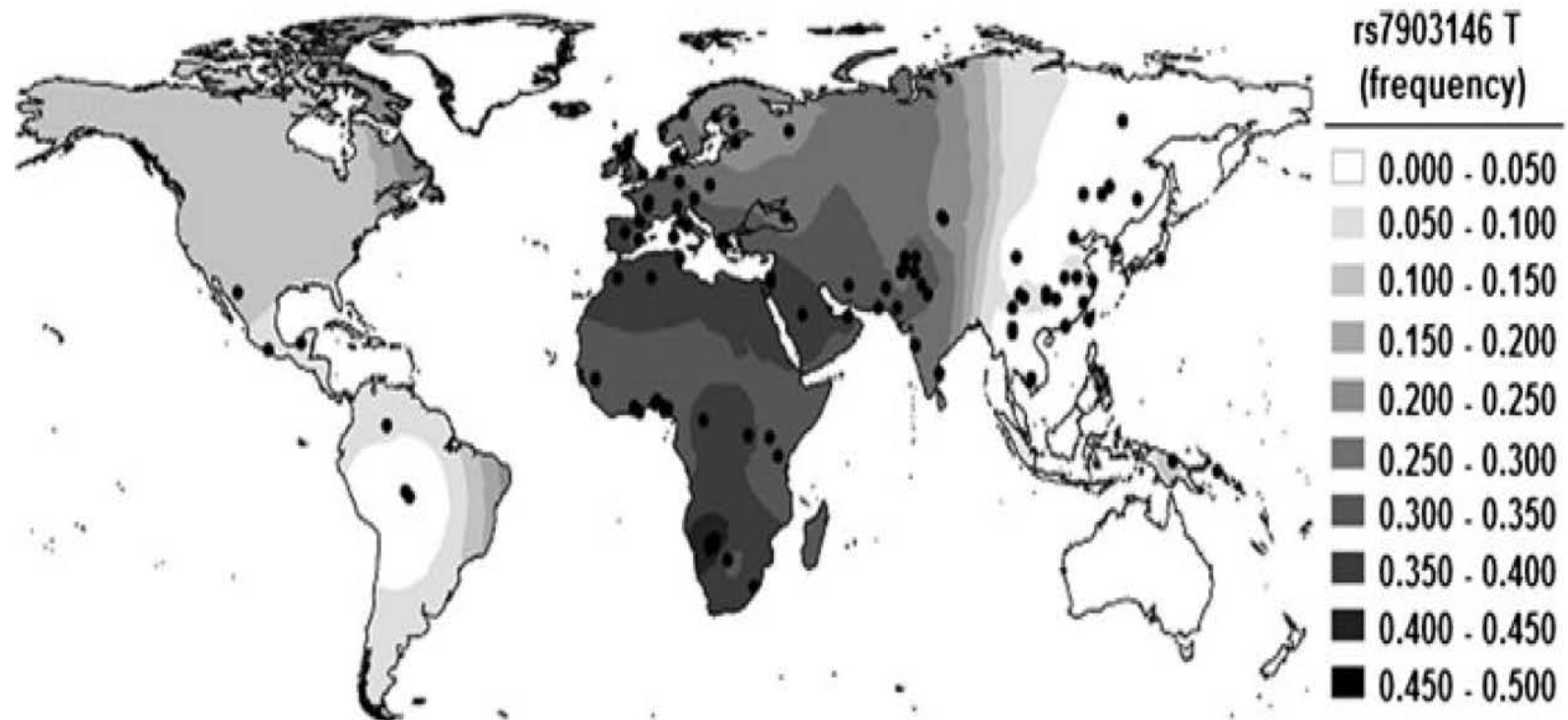
Source: Langenberg et al, PLoS Med 2014

# Phase 3: Moving from within-population investigation to the study of between-population differences



# Studying between-population differences – genetics

Global distribution of rs7903146 T allele in TCF7L2

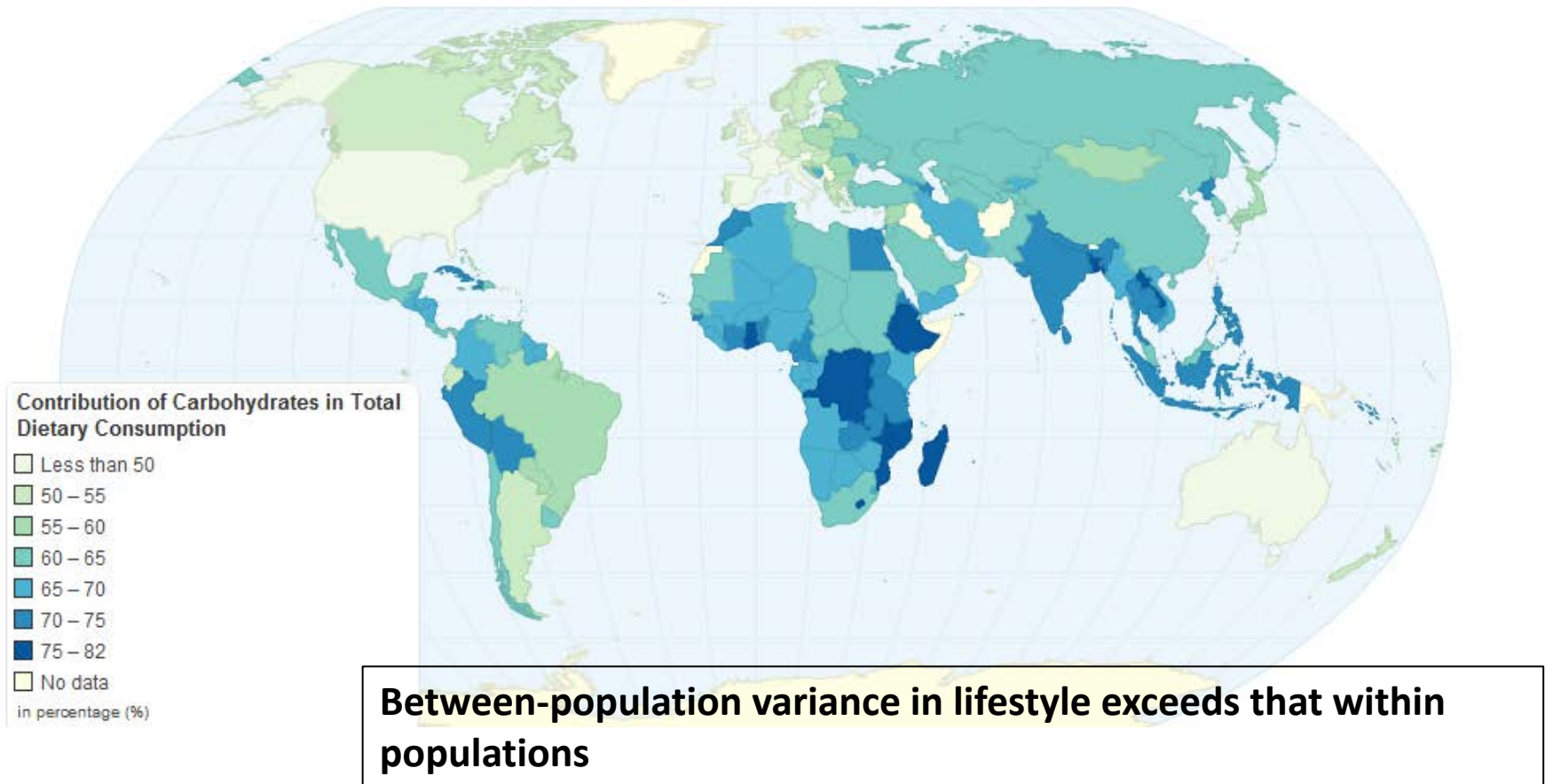


Source: Guinan, Biochem Genet 2012

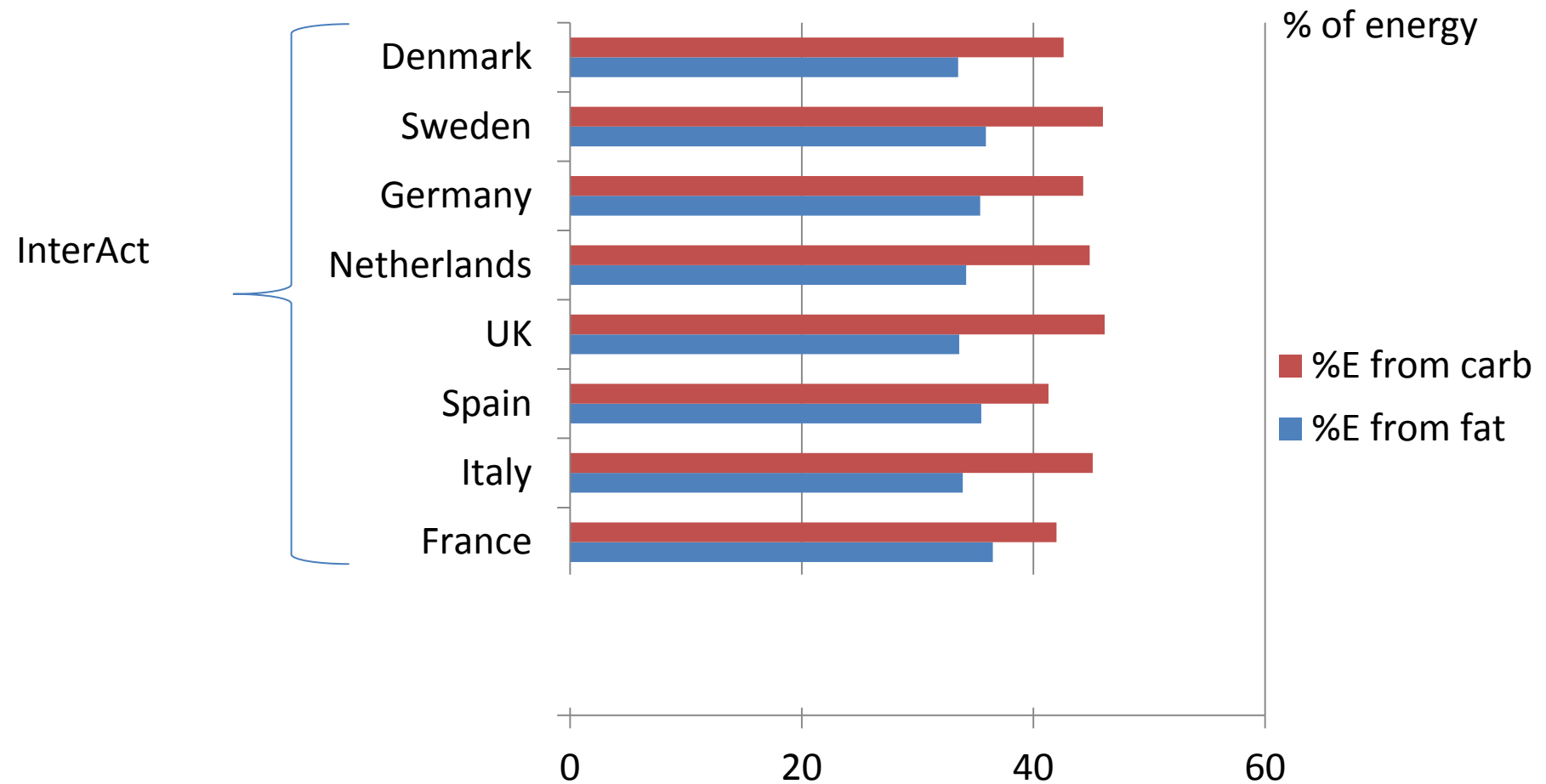


# Global variation in carbohydrate intake

## Contribution of Carbohydrates in Total Dietary Consumption



# Percentage energy (%E) from fat and carbohydrates



Source: Nanri et al, Am J Clin Nutr, 2011

# How to realise the vision of bringing data together to allow the study of between-population differences in risk

---

- Find relevant studies globally
- Find out what data the studies have collected
- Find an appropriate way of bringing data together
- Find a way of interpreting different forms of data that are brought together

# Programme

---

<b>14:30</b>	<b>Welcome</b>	Nick Wareham
<b>14:35</b>	<b>Scientific opportunity, challenge and vision</b>	
14:35	Understanding differences in risk of diabetes and obesity between populations	Nick Wareham
14:50	<b><i>Challenges of data sharing models</i></b>	<b><i>Nita Forouhi</i></b>
15:10	InterConnect: Vision of a changed paradigm	Nick Wareham
<b>15:30</b>	<b>Delivering the InterConnect vision</b>	
15:30	Data discovery: The registry	Matthias Schulze
15:45	Bringing the analysis to the data: Proof of concept	Tom Bishop
16:00	Developing the vision via exemplar research questions	Ken Ong
<b>16:15</b>	<b>Open discussion and involvement</b>	Nick Wareham



**Inter  
Connect**



*Global data for diabetes and obesity research*

# Challenges of data sharing models

***Nita Forouhi***

***InterConnect WP4 Leader & MRC Epidemiology Unit, University of Cambridge, UK***

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

# Data sharing models

---

Consider the models about how data is currently shared, and might be shared in the future

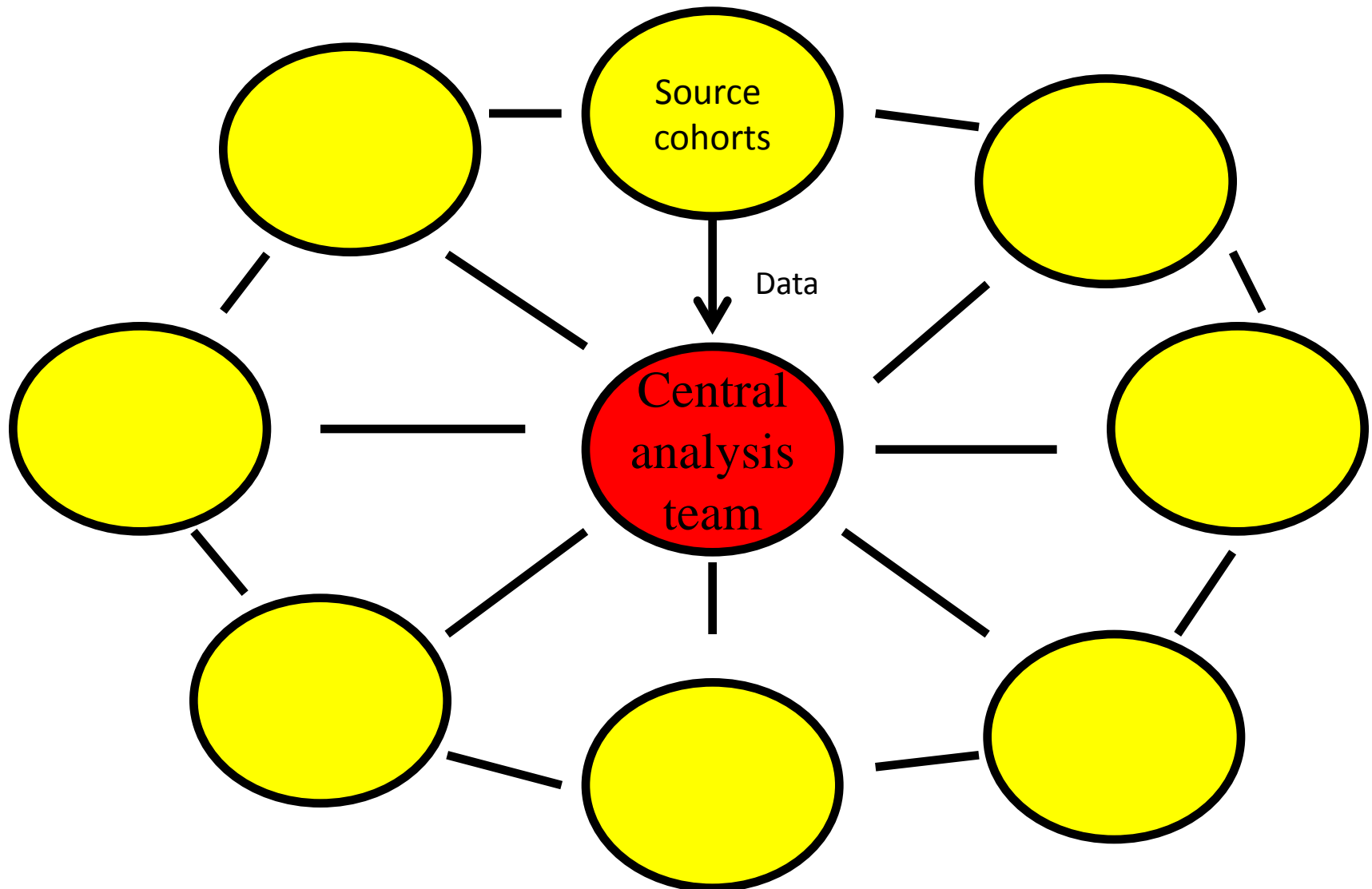
Consider:

- Possible benefits and difficulties of each model
- Issues from different perspectives – i.e that of a researcher, a funder, etc

Think of a future world in which we are trying to connect multiple studies together across different countries

# 1. Sharing of data between cohorts using traditional collaboration/consortia agreements

---



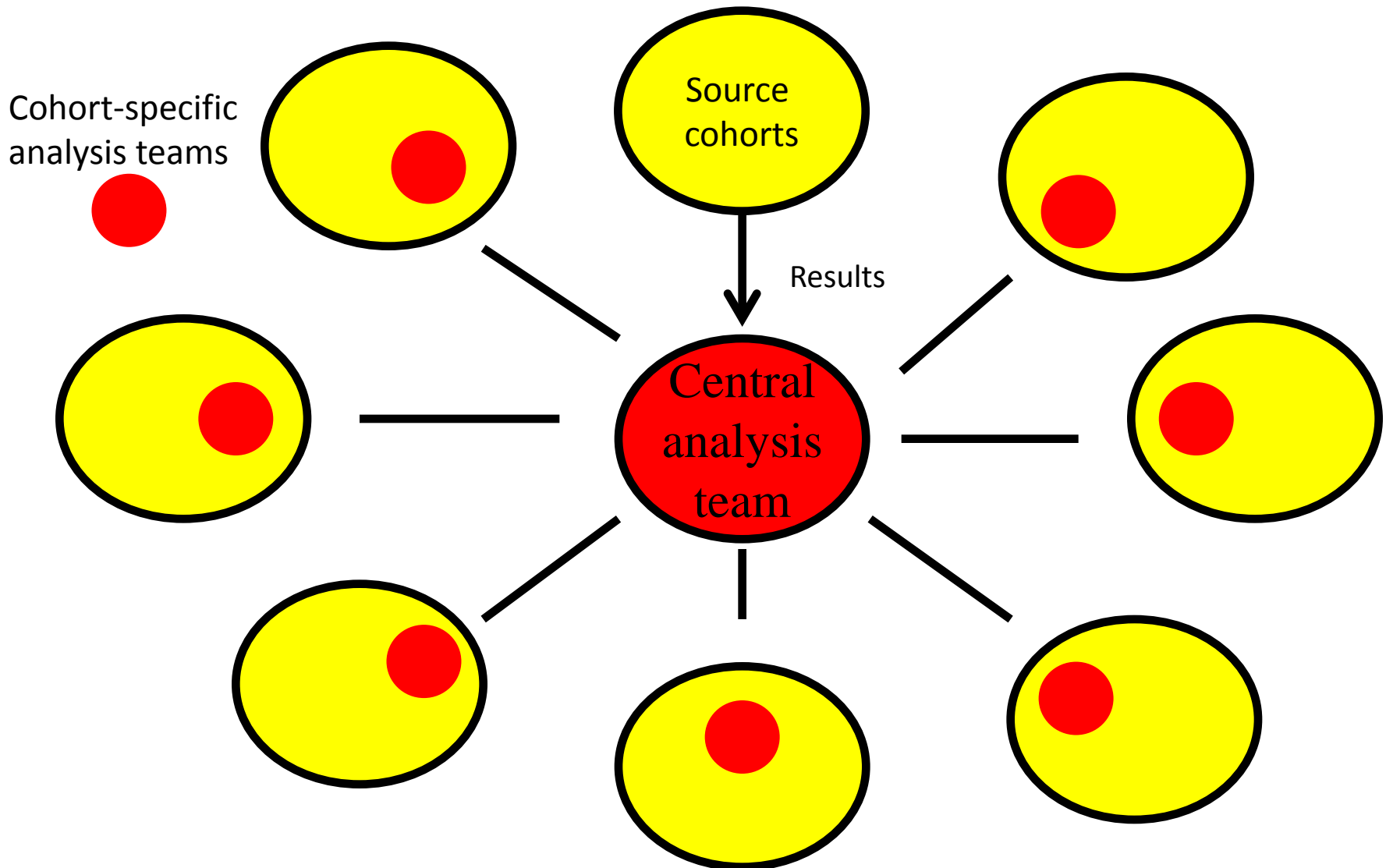
# Possible issues: Model 1- sharing of data

Benefits	Challenges
Enables physical sharing of individual level data	<p>Considerable transactional burden</p> <ul style="list-style-type: none"><li>• Burden will increase exponentially as number of partners in consortia increases</li><li>• Contracts</li><li>• Regulatory processes, e.g. cross border transfer</li><li>• Data transfer problems and diversity of attitudes can be limiting</li><li>• Need well established collaborative networks between partners – lengthy process, requires trust</li><li>• Bringing in a global perspective will add substantially to the complexity</li></ul>
Enables in-depth individual level meta analysis	<p>Difficult to control passage of data and use beyond the original intention</p> <p>If centralised around a sole analytical centre, resentment may arise about imbalance of opportunities to lead as opposed to contribute</p>



## 2. Ad hoc consortia - sharing of results

---



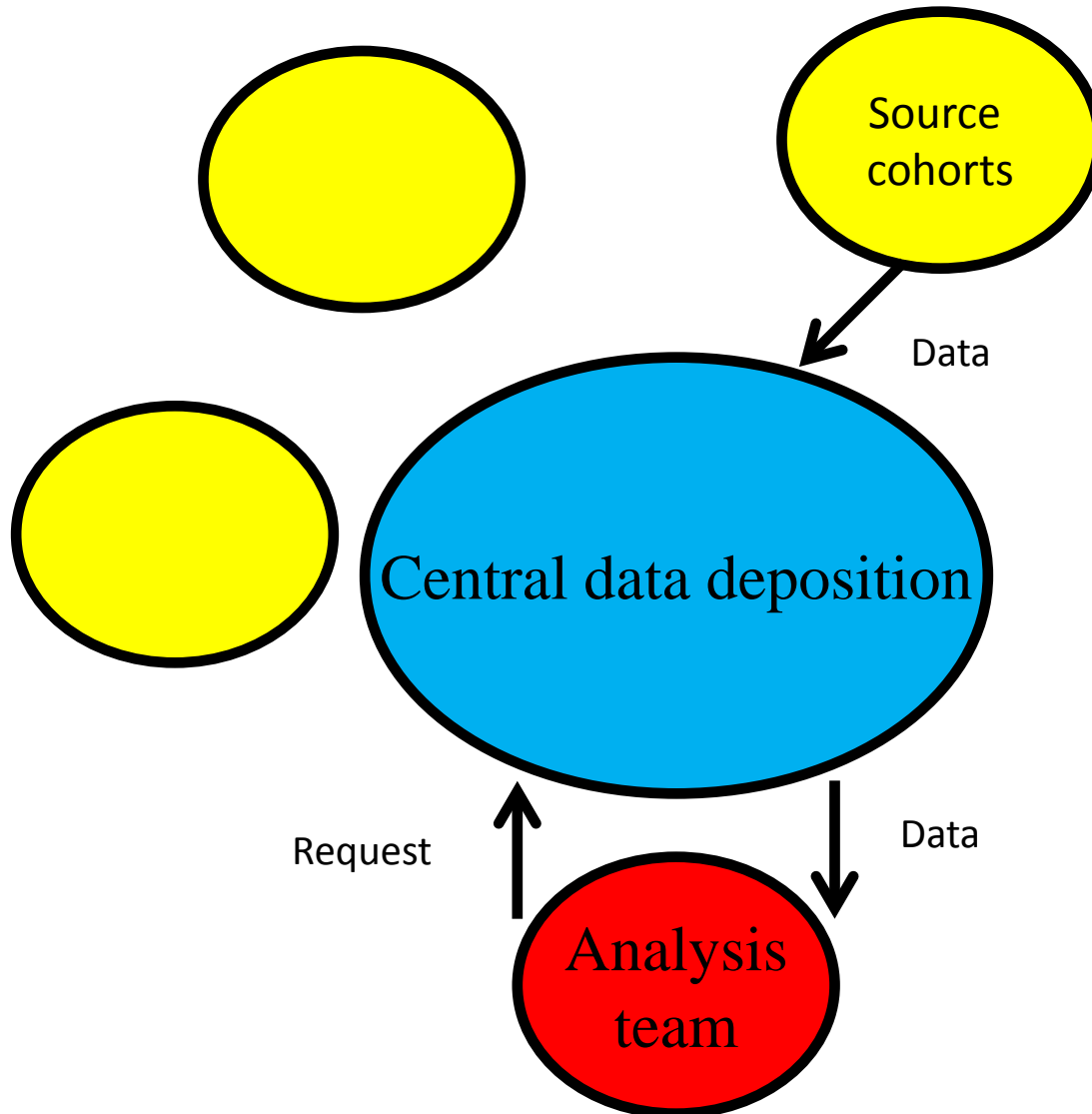
# Possible issues: : Model 2- sharing of results

---

Benefits	Challenges
Ad hoc consortia work well for genetic analyses, allowing sharing of RESULTS without administrative or organisational complexity	<p>Limits of analysis</p> <ul style="list-style-type: none"><li>• When results are meta-analysed rather than data, important details may be missed when analysed across populations</li><li>• Limits of meta-analysing interaction terms from individual studies</li><li>• Difficulties of data harmonisation given limited attention</li><li>• Analysis potentially misses major between-cohort variation</li></ul>
Some ethical/legal issues are eased	<p>Each cohort/centre needs analytical capacity</p> <ul style="list-style-type: none"><li>• Each centre may be inundated with large number of requests</li><li>• Analytical effort is decentralised to individual studies who spend a massive amount of time servicing the work of others</li><li>• This is time consuming for investigators, and may be of concern to funders</li></ul>

### 3. Central deposition of data

---



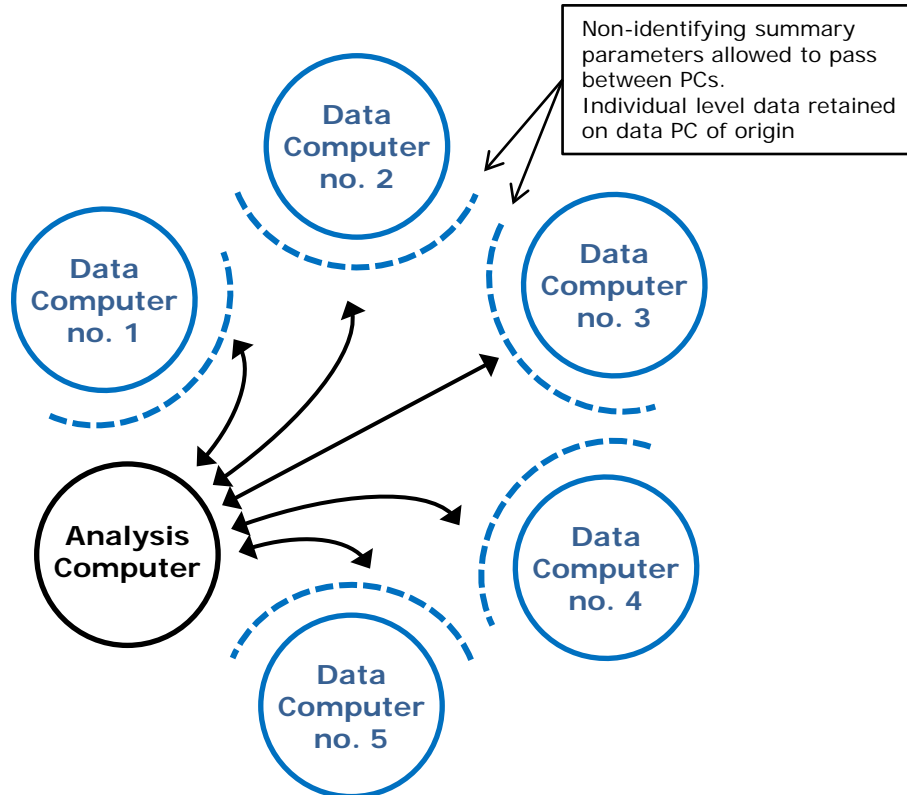
# Possible issues: Model 3- deposit data centrally

---

Benefits	Challenges
Approach works within some countries for some forms of data	Likelihood of success for between-country collaboration low e.g. access decisions need delegated authority; substantial challenge on a global scale
Can provide greater opportunity to wide range of researchers to access the data	Major governance, ethical and legal challenges e.g. who owns the data Unlikely to work for more complex forms of data Difficult to mandate for historical data

# The future: Federated meta-analysis

---

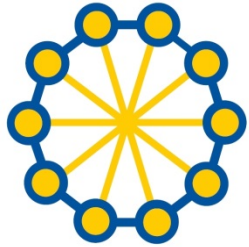


- Data stays within governance structure of source cohort
- Cohorts focus efforts on preparation of data and IT infrastructure for sharing
- Analytical effort more focused on the scientific –led questions

# Programme

---

<b>14:30</b>	<b>Welcome</b>	Nick Wareham
<b>14:35</b>	<b>Scientific opportunity, challenge and vision</b>	
14:35	Understanding differences in risk of diabetes and obesity between populations	Nick Wareham
14:50	Challenges of data sharing models	Nita Forouhi
15:10	<b><i>InterConnect: Vision of a changed paradigm</i></b>	<b><i>Nick Wareham</i></b>
<b>15:30</b>	<b>Delivering the InterConnect vision</b>	
15:30	Data discovery: The registry	Matthias Schulze
15:45	Bringing the analysis to the data: Proof of concept	Tom Bishop
16:00	Developing the vision via exemplar research questions	Ken Ong
<b>16:15</b>	<b>Open discussion and involvement</b>	Nick Wareham



**Inter  
Connect**



*Global data for diabetes and obesity research*

# **InterConnect: Vision of a changed paradigm**

*Nick Wareham*

*InterConnect Co-ordinator & Director, MRC Epidemiology Unit,  
University of Cambridge, UK*

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

# InterConnect vision

---

- Goal to optimise the use of existing data to enable cross-cohort analyses



# Barriers to cross-cohort analyses

---

## Results sharing:

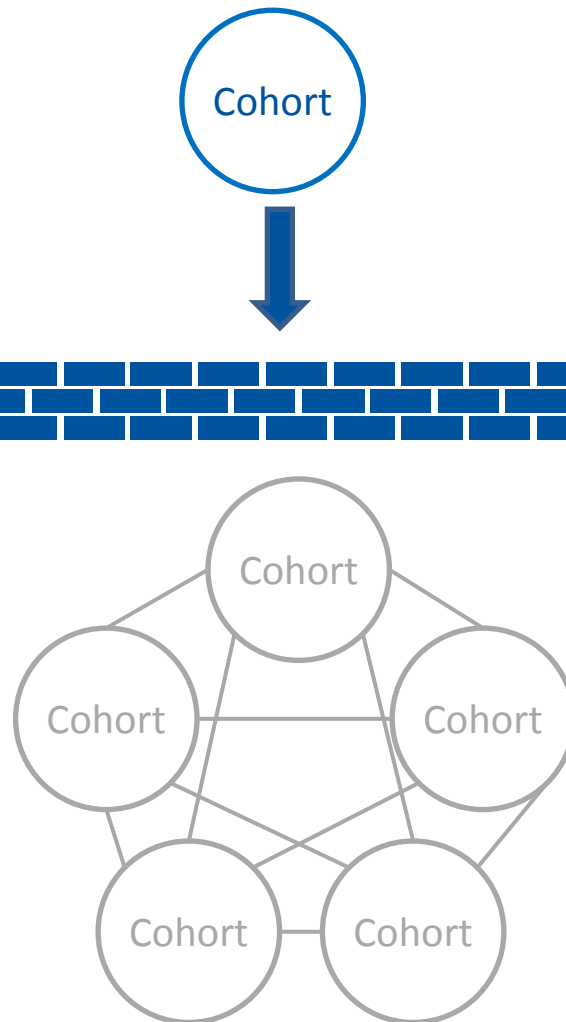
Burden on collaborators of repeatedly preparing and analysing data

Results sharing works well for some risk factors but misses between cohort variation for others

## Data pooling:

Collaborators fear loss of ownership of their data

Complex data-sharing or deposition agreements are needed



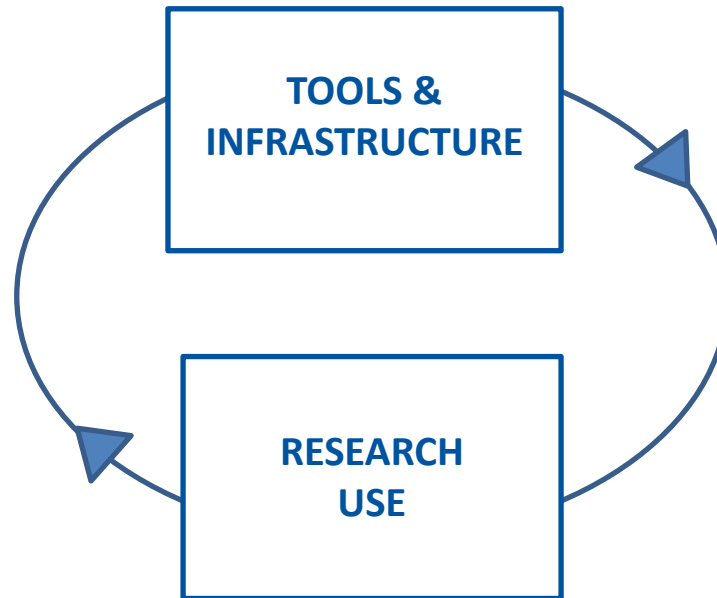
# InterConnect vision

---

- Goal to optimise use of existing data to enable cross-cohort analyses
  - Individual participant meta-analysis of pooled data from separate cohorts is analytically desirable
  - InterConnect aims to enable a solution without physical pooling of data by **TAKING THE ANALYSIS TO THE DATA** through federated meta-analysis

# InterConnect: A bridging function

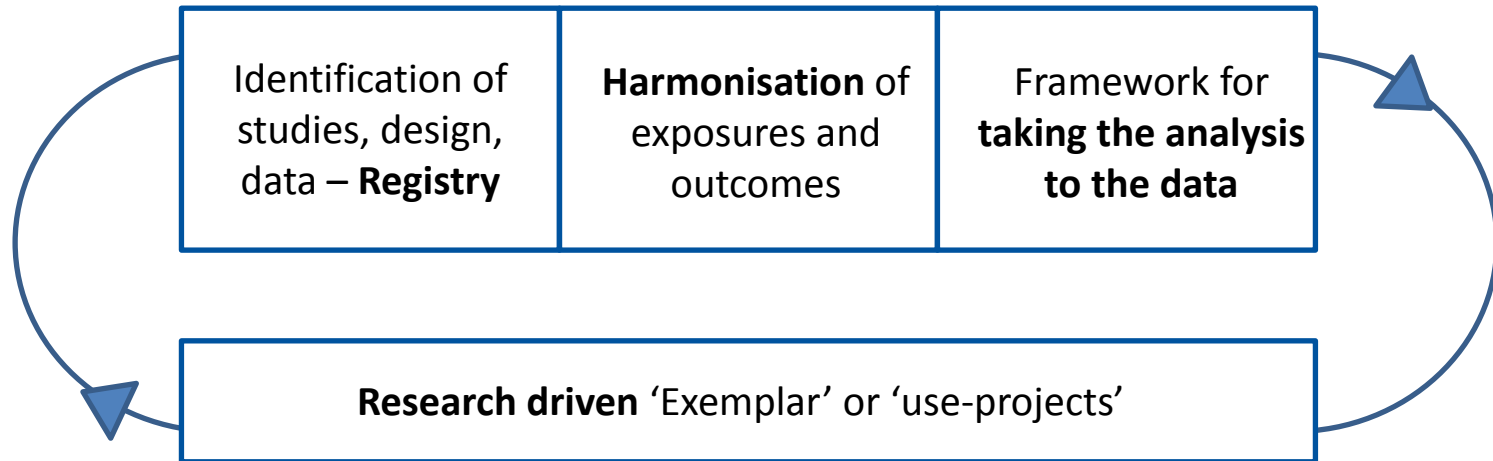
---



# InterConnect: A bridging function




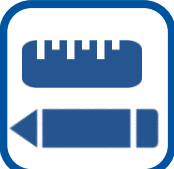
---

## TOOLS & INFRASTRUCTURE



## RESEARCH USE: APPLICATION TO FOCUS & REFINE

<p>Identification of studies, design, data – <b>Registry</b></p>	<p><b>Harmonisation</b> of exposures and outcomes</p>	<p>Framework for <b>taking the analysis to the data</b></p>
--	---	---

	<p>A catalogue of studies relating to diabetes and obesity</p>
	<p>Populations recruited to the study</p>
	<p>Biological samples stored or analysed</p>
	<p>The study design that was employed</p>

Identification of studies, design, data – **Registry**

**Harmonisation** of exposures and outcomes

Framework for **taking the analysis to the data**

**Exemplar question: Study A**

In a typical week, how many glasses of red wine (6 ounces) do you drink per day?

[\_\_\_] Number of drinks per day

**Exemplar question: Study B**

In general, how many glasses of red wine do you drink per day over a week and weekend?

Week: [\_\_\_] Number/day

Weekend: [\_\_\_] Number/day

**Exemplar question: Study C**

In a typical week, how many glasses of red wine do you drink per day?

1–3

4–6

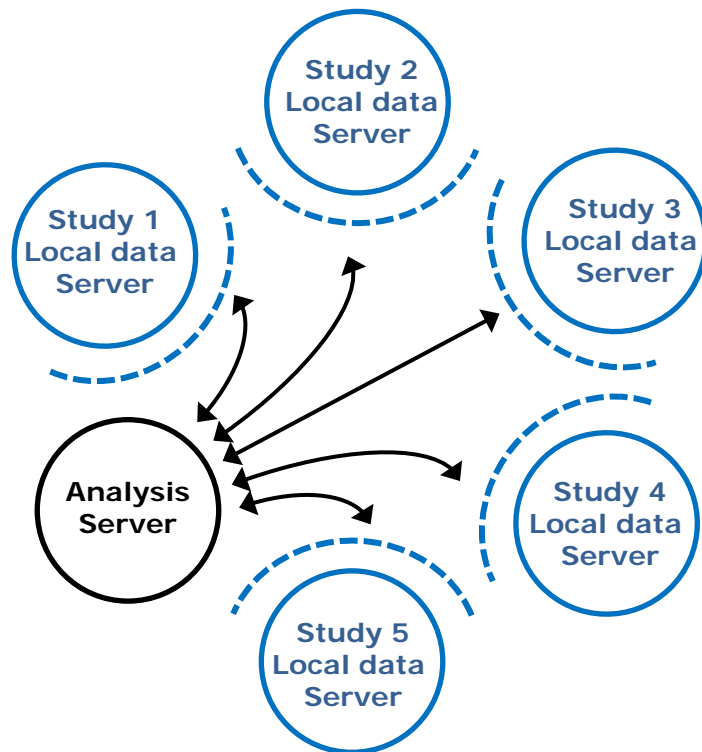
7–9

10 or more

Align to give a single exposure where possible

InterConnect software captures how the alignment is made so that it is both explicit and re-usable

Identification of studies, design, data – <b>Registry</b>	<b>Harmonisation</b> of exposures and outcomes	Framework for <b>taking the analysis to the data</b>
---	--	--

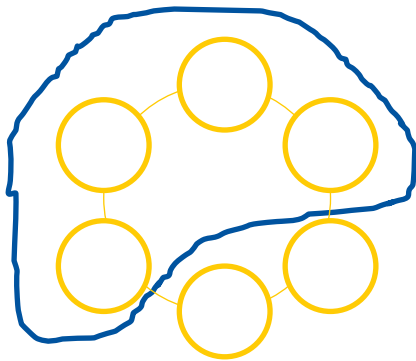


- Data stay within the governance structure of the cohort
- Analytical instructions and non-identifying summary parameters allowed to pass between computers
- Any user with appropriate log in credentials can remotely access the analysis server to run analysis code

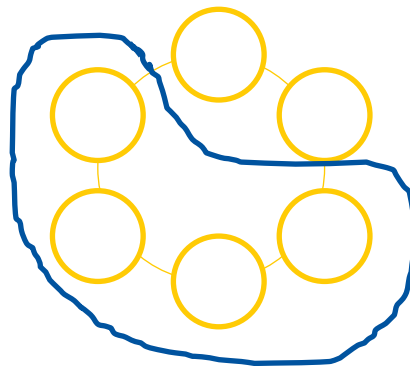
# Vision - a dynamic network

---

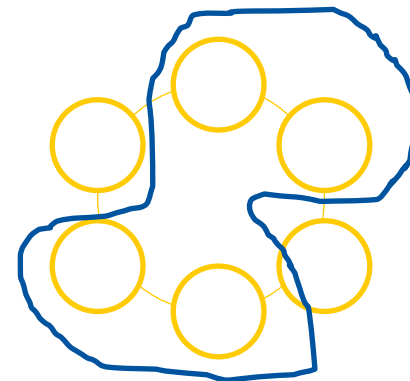
- InterConnect is NOT an analytical consortium
  - Enabling ad hoc consortia to form to answer questions that require cross-cohort analysis
  - Cohorts join network and decide what research to participate in



Consortium 1  
Question A



Consortium 2  
Question B

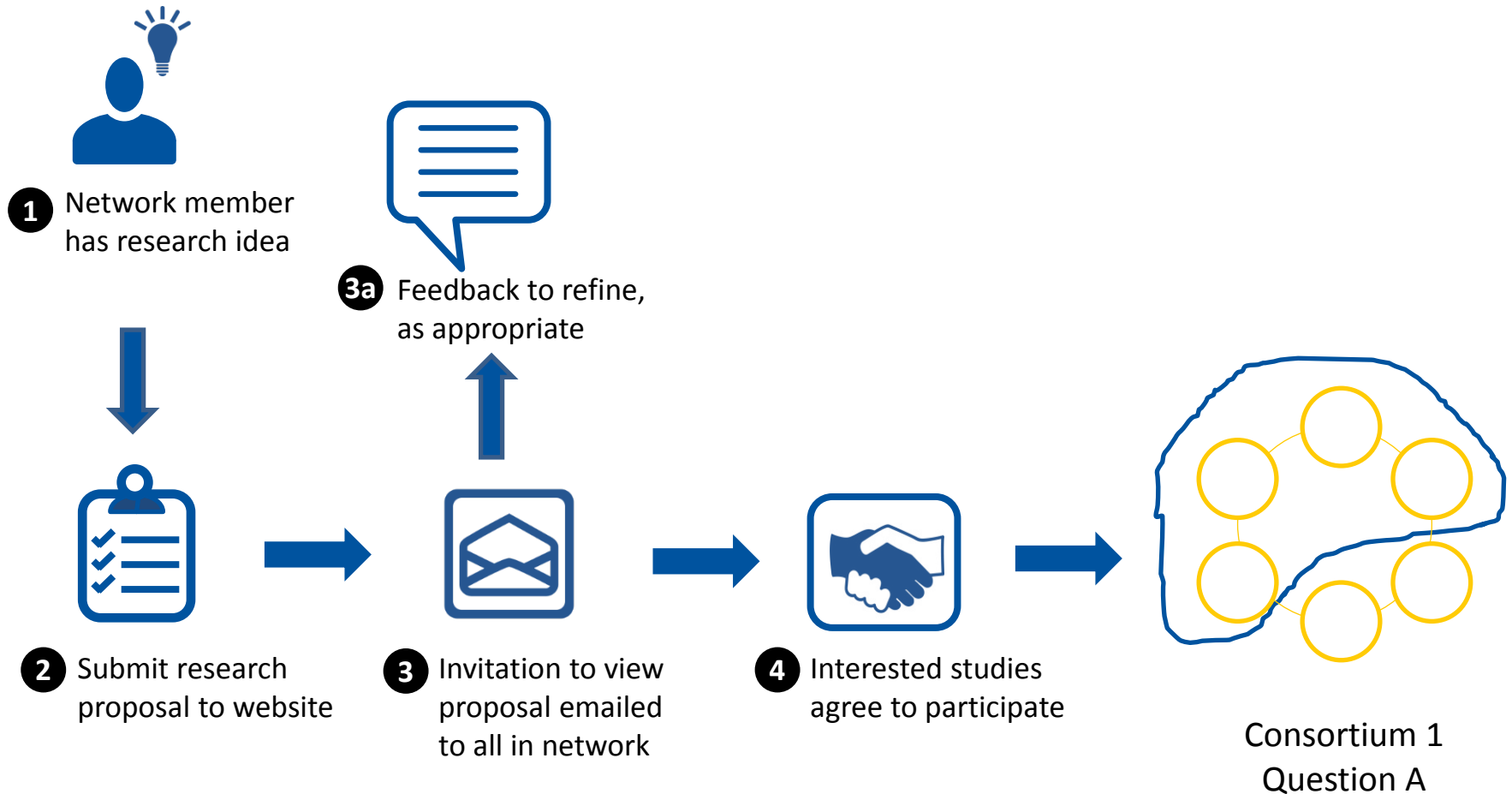


Consortium 3  
Question C



# How will consortia form?

---



# Who decides the rules?

---

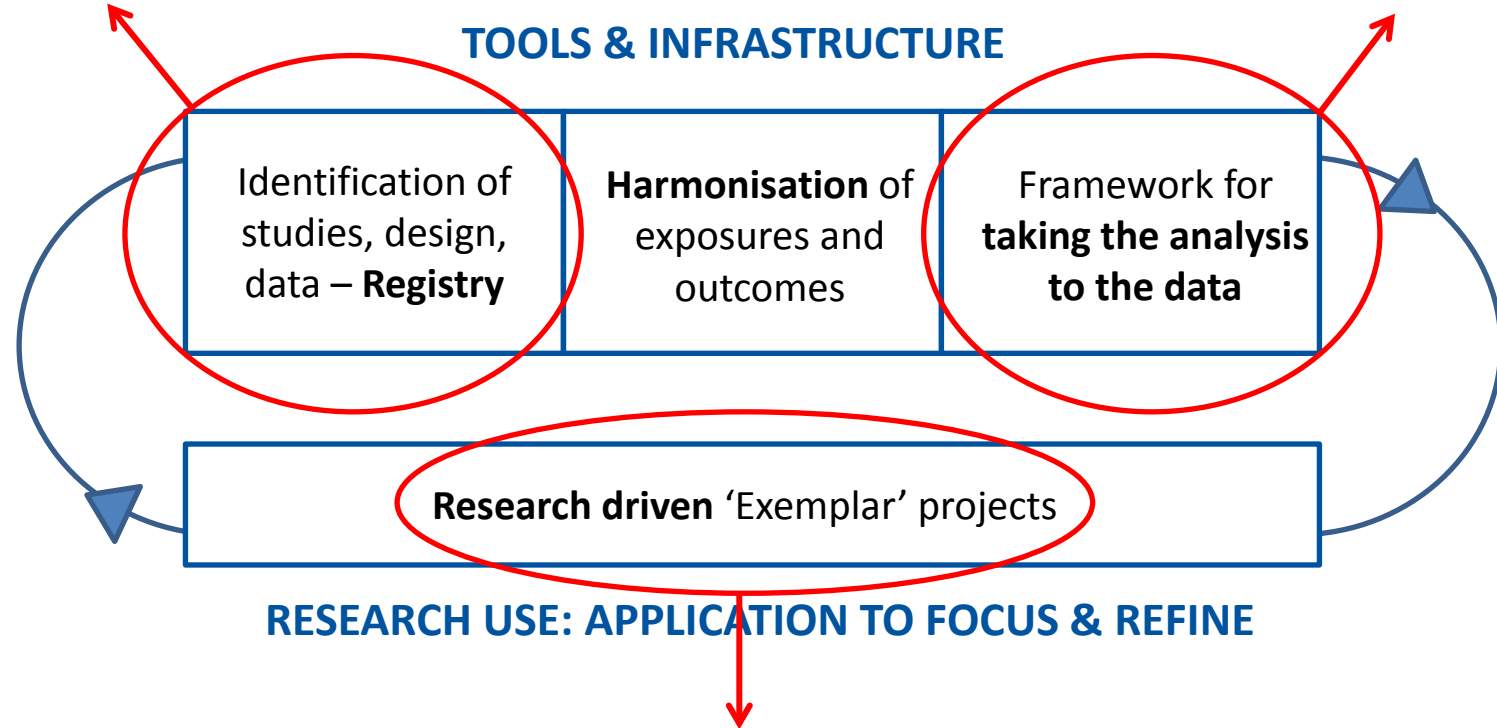
- Each ad hoc consortium will decide its own way of working and be autonomous

# Session 2: Delivering the vision

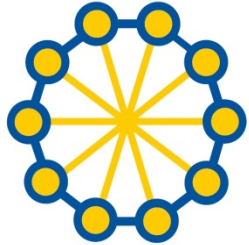
---

*Registry now live & developing*

*Proof of concept – federated approach*



*Forming ad hoc consortia – PA in pregnancy*



**Inter**  
**Connect**



*Global data for diabetes and obesity research*

## Acknowledgement

- This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

## Connect with us

- [InterConnect@mrc-epid.cam.ac.uk](mailto:InterConnect@mrc-epid.cam.ac.uk)
- [www.interconnect-diabetes.eu](http://www.interconnect-diabetes.eu)



*Global data for diabetes and obesity research*

## **Data discovery: The registry**

***Matthias Schulze***

***InterConnect WP1 Leader & German Institute of Human Nutrition  
Potsdam-Rehbrücke, Germany***

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.





Identification of studies, design, data – **Registry**

**Harmonisation** of exposures and outcomes

Framework for **taking the analysis to the data**



## A catalogue of studies relating to diabetes and obesity

	Populations recruited to the study
	The study design that was employed
	Data which have been collected
	Biological samples stored or analysed

# Developing a study registry

---

- Tasks of the InterConnect project
  - Setup a database to include information about studies
  - Prepare a standardised web-based procedure for data input for project partners and external investigators
  - Prepare a registry website which hosts the visualization of the registry database

# 2-Phase registry

---

- Phase 1: “broad and shallow”
  - Simple but useful information
  - Largely collected based on available/public information
  
- Phase 2: in depth information
  - To be collected directly from studies



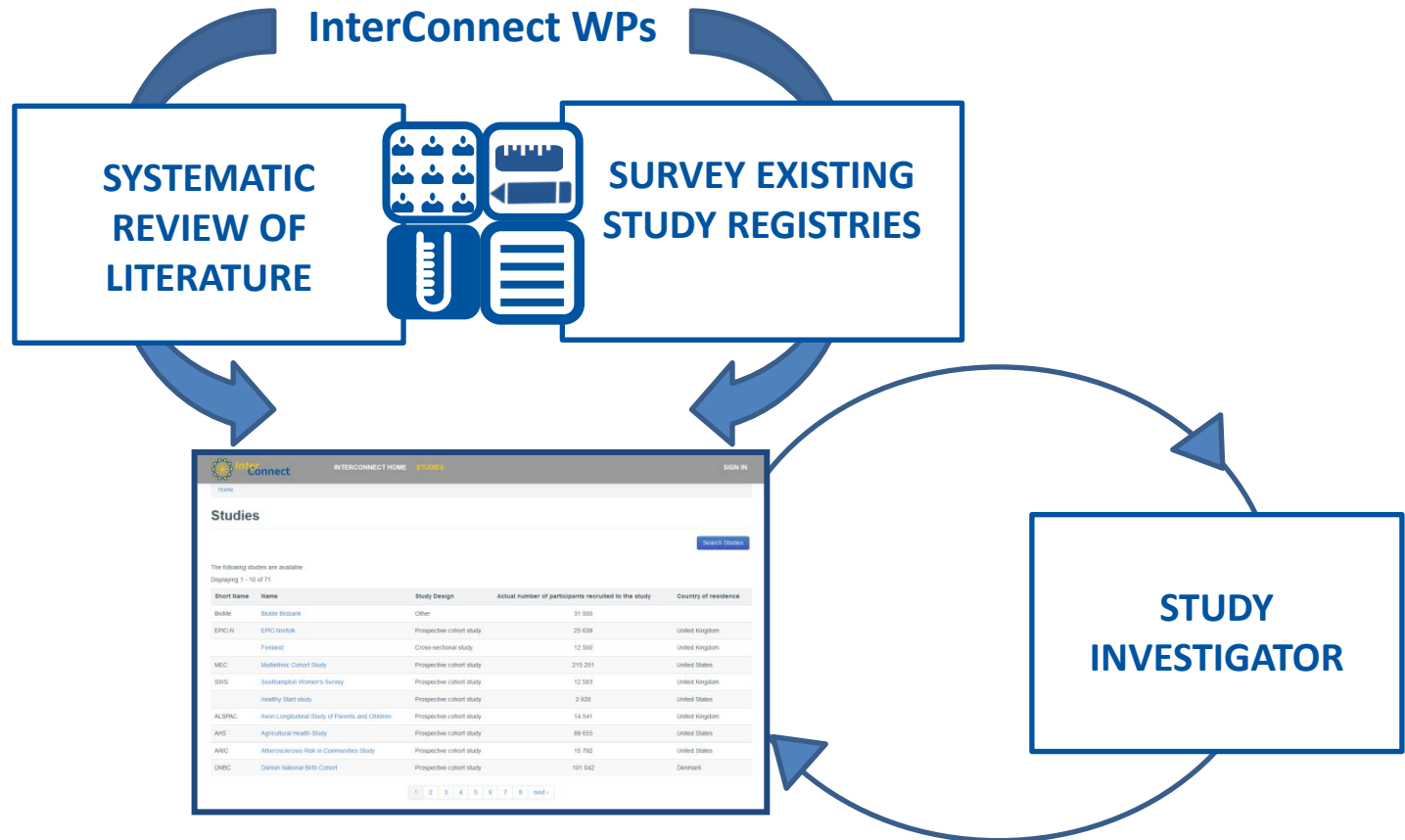
# Phase 1 information

---

- General information (study name, contact persons, web link)
- Study design
- Ethnicity and race
- Sampling frame and recruitment target
- Health information collected at baseline/follow-up
- Key exposures (diet, activity, DNA sources)

# Phase 1 information

- Phase 1 information



# Web-based data input

The screenshot shows a web browser window with the URL `https://studies.interconnect-diabetes.eu/node/add/study`. The browser's address bar and tabs are visible at the top. Below the browser window, the website's header includes the 'InterConnect' logo, navigation links for 'INTERCONNECT HOME' and 'STUDIES', and a 'USER MENU' button. The main content area features a breadcrumb trail 'Home / Add content' and a heading 'Create Study'. A yellow notification bar states: 'New content: Your draft will be placed in moderation.' Below this, a tabbed interface is shown with the 'General Information' tab selected. The form contains several sections: 'Name of the study' with a 'Name \*' field and an 'Official study name' input; 'Acronym' with an 'Acronym of the study if applicable.' input; 'Contacts' with a table for adding contact names and a '+ Create Contact' button; and 'Website' with 'Title' and 'URL' input fields. A blue button labeled 'Add another item' is positioned above the website section.

Home / Add content

## Create Study

New content: Your draft will be placed in moderation.

**General Information \*** | Methods \* | Ethnic and racial groups recruited | Available Information | Populations from which the sample(s) is drawn

**Name of the study**

**Name \***  
Official study name

**Acronym**  
Acronym of the study if applicable.

**Contacts**  
Enter the Contact full name. You will be able to edit these Contacts once the Study will be saved. [Show row weights](#)

<input type="text"/>
----------------------

+ Create Contact

[Add another item](#)

**Website**

Title	URL
<input type="text"/>	<input type="text"/>

# Web-based data input



New content: Your draft will be placed in moderation.

- General Information \*
- Methods \*
- Ethnic and racial groups recruited
- Available Information
- Populations from which the sample(s) is drawn

## HEALTH INFORMATION

- Anthropometric traits (e.g. BMI, waist circumference) at baseline
- Anthropometric traits (e.g. BMI, waist circumference) during follow up
- Glycaemic traits (e.g. glucose, insulin) at baseline
- Glycaemic traits (e.g. glucose, insulin) during follow up
- Prevalent type 1 diabetes at baseline
- Incident type 1 diabetes during follow up
- Prevalent type 2 diabetes at baseline
- Incident type 2 diabetes during follow up
- History of gestational diabetes
- Incident gestational diabetes during follow up
- Other types of diabetes prevalent at baseline
- Other types of diabetes incident during follow up



**OBESITY  
DIABETES**

## KEY EXPOSURES

### Dietary measures

Does the study have any measures of diet?

Yes

No

# The InterConnect study registry online

<https://studies.interconnect-diabetes.eu/studies>

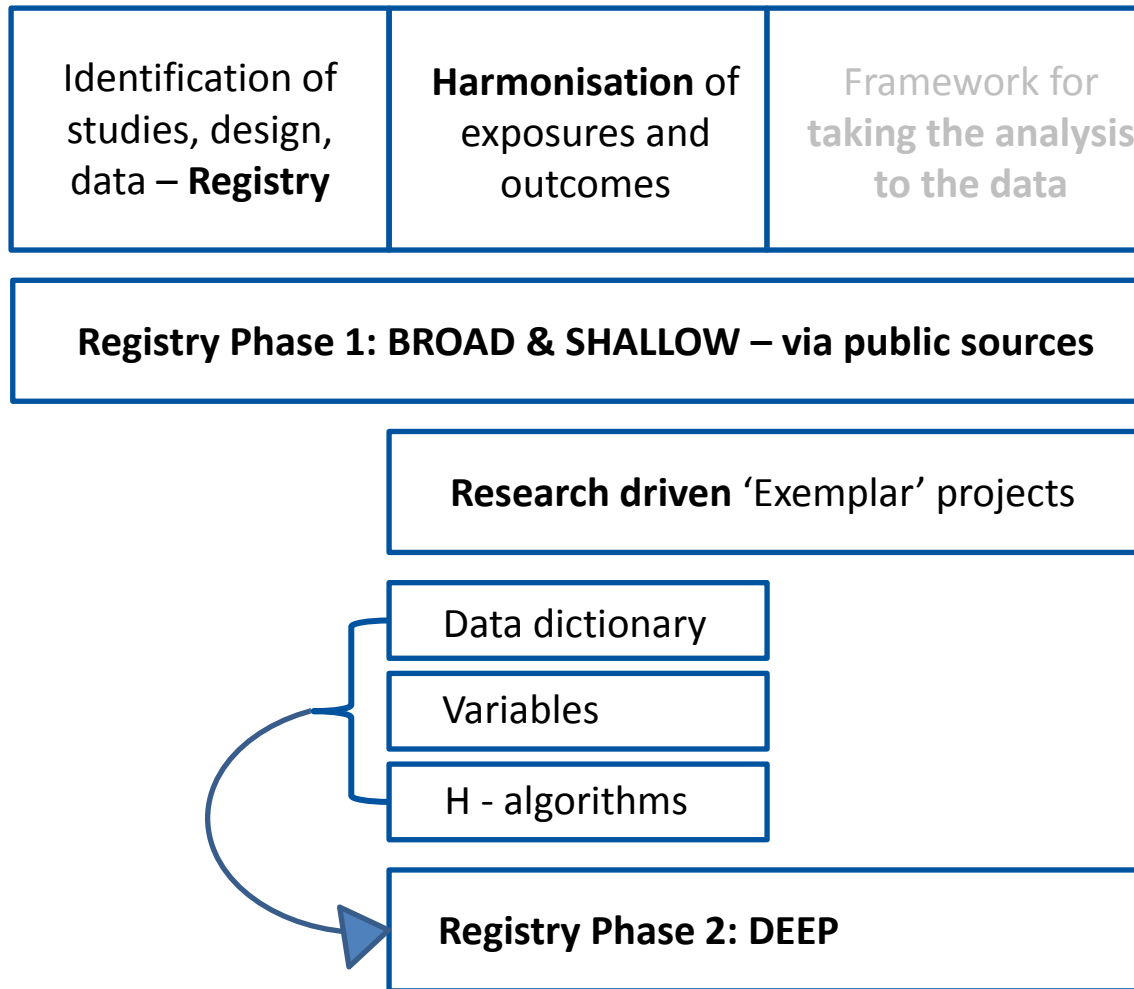
The screenshot shows the InterConnect website interface. At the top, there is a navigation bar with the InterConnect logo, "INTERCONNECT HOME", "STUDIES", and "SIGN IN". Below the navigation bar, there is a "Home" link and a "Studies" heading. A "Search Studies" button is located on the right side. The main content area displays a list of studies with the following columns: Short Name, Name, Study Design, Actual number of participants recruited to the study, and Country of residence. The list includes studies such as BioMe, EPIC-N, MEC, SWS, ALSPAC, AHS, ARIC, and DNBC. A summary box titled "Studies in Registry (as of Sept. 10, 2015)" is overlaid on the right side of the list, showing the following counts: Verified (25), Public (46), and Total in progress (81). A pagination bar at the bottom of the list shows page numbers 1 through 8 and a "next" button.

Short Name	Name	Study Design	Actual number of participants recruited to the study	Country of residence
BioMe	BioMe Biobank	Other	31 000	
EPIC-N	EPIC-Norfolk	Prospective co		
	Fenland	Cross-sectiona		
MEC	Multiethnic Cohort Study	Prospective co		
SWS	Southampton Women's Survey	Prospective co		
	Healthy Start study	Prospective co		
ALSPAC	Avon Longitudinal Study of Parents and Children	Prospective co		
AHS	Agricultural Health Study	Prospective co		
ARIC	Atherosclerosis Risk in Communities Study	Prospective cohort study	15 792	United States
DNBC	Danish National Birth Cohort	Prospective cohort study	101 042	Denmark

Verified	Public	Total in progress
25	46	81

# Online, re-usable resource: Registry phase 2

---



# Illustrations – BioSHaRE, Maelstrom Research

- Record potential to re-use harmonised variables across studies

## Harmonization

**⊖ Undetermined** - the harmonization potential of this variable has not yet been evaluated.

**✓ Complete** - the study assessment item(s) (e.g. survey question, physical measure, biochemical measure) allow construction of the variable as defined in the dataset.

**✗ Impossible** - there is no information or insufficient information collected by this study to allow the construction of the variable as defined in the dataset.

Download

Variable	Atlantic PATH 1	Atlantic PATH 2	BCGP 1	BCGP 2	BCGP 3	CaG	OHS 1	OHS 2	TTP 1	TTP 2
S_DIS_EMPHYSEMA_AGE	✗	✗	✗	✗	✓	✓	✗	✓	✓	✗
S_DIS_EMPHYSEMA_TX	✗	✗	✗	✗	✓	✓	✗	✗	✓	✗
S_DIS_CB_EVER	✗	✗	✗	✗	✓	✓	✗	✓	✓	✗
S_DIS_CB_AGE	✗	✗	✗	✗	✓	✓	✗	✓	✓	✗
S_DIS_CB_TX	✗	✗	✗	✗	✓	✓	✗	✗	✓	✗
S_DIS_COPD_EVER	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓
S_DIS_COPD_AGE	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓
S_DIS_DEP_EVER	✓	✓	✓	✓	✗	✗	✓	✓	✗	✓
S_DIS_DEP_AGE	✓	✓	✓	✓	✗	✗	✓	✓	✗	✓
A_DIS_DIAB_EVER	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

previous 1 2 3 4 5 6 7 8 9 ... next

# Illustrations – BioSHaRE, Maelstrom Research

- *Summary statistics for harmonised variables across studies*

## Harmonization

Download

● **Undetermined** - the harmonization potential of this variable has not yet been evaluated.  
✓ **Complete** - the study assessment item(s) (e.g. survey question, physical measure, biochemical measure) allow construction of the variable as defined in the dataset.  
✗ **Impossible** - there is no information or insufficient information collected by this study to allow the construction of the variable as defined in the datasets.

Variable	Atlantic PATH 1	Atlantic PATH 2	BCGP 1	BCGP 2	BCGP 3	CaG	OHS 1	OHS 2	TTP 1	TTP 2
S_DIS_EMPHYSEMA_AGE	✗	✗	✗	✗	✓	✓	✗	✓	✓	✗
S_DIS_EMPHYSEMA_TX	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
S_DIS_CB_EVER	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
S_DIS_CB_AGE	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
S_DIS_CB_TX	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
S_DIS_COPD_EVER	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
S_DIS_COPD_AGE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
S_DIS_DEP_EVER	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
S_DIS_DEP_AGE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
A_DIS_DIAB_EVER	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Age of the participant in years (continuous).

**Description**

**Label:**  
Age in Years (continuous)

**Dataset:**  
[Healthy Obese Project DataSchema](#)

**Value Type:**  
Integer

**Unit:**  
Years

**Repeatable:**  
No

**Domains**

**Data Source:**  
Questionnaire

**Sociodemographic/Socioeconomic Characteristics:**  
Age/Birth date

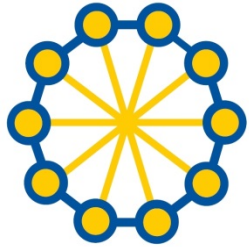
**Statistics**

Study	Min	Max	Mean	Std. Dev	Count
HUNT	19.000	101.000	49.682	17.244	65241
LifeLines	18.000	94.000	45.136	12.516	94516
KORA	32.000	83.000	56.678	13.229	3080
PREVEND	29.000	75.000	49.747	12.698	8592
NCDS	44.000	46.000	44.850	0.464	7210
FINRISK 2007	25.000	74.000	52.619	13.519	5024

**Categories**

Name	Label	Missing
999	Missing	✓





**Inter**  
**Connect**



*Global data for diabetes and obesity research*

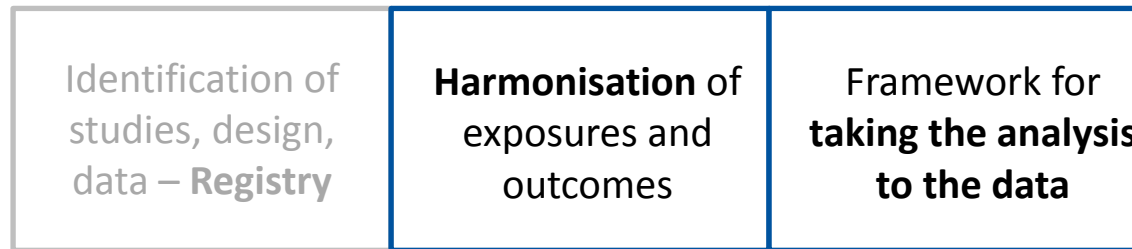
# **Bringing the analysis to the data: Proof of Concept**

*Tom Bishop,  
Technical Lead, MRC Epidemiology Unit, University of  
Cambridge, UK*

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

# Bringing the analysis to the data

---



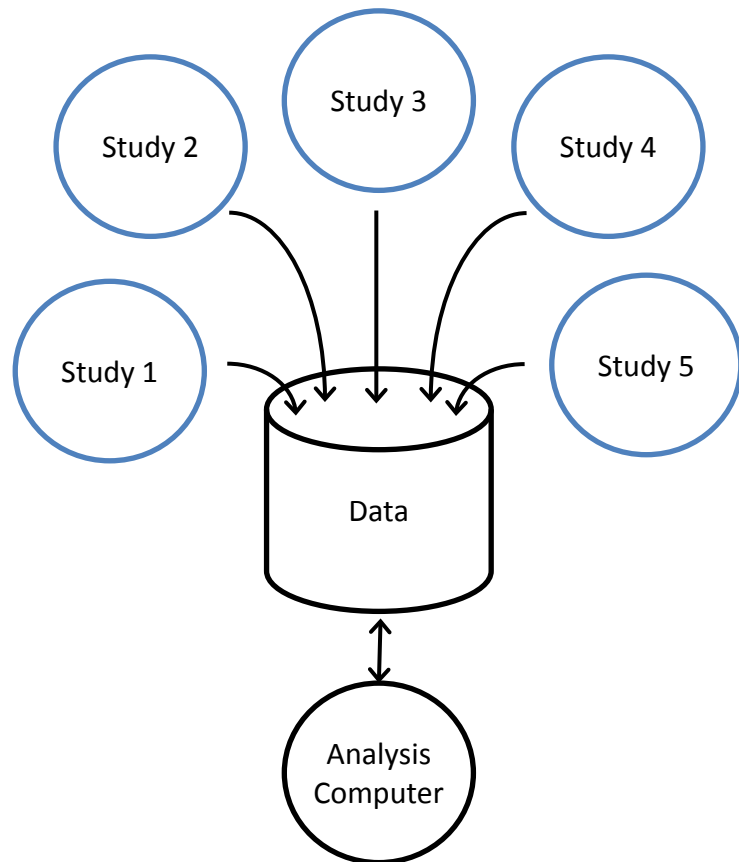
Proof of concept to show that the InterConnect technologies and methods work:

- Harmonisation and federated analysis
- Description of technical requirements to join InterConnect
- Overview of InterConnect security features

# Proof of concept: test technology for harmonisation and federated analysis

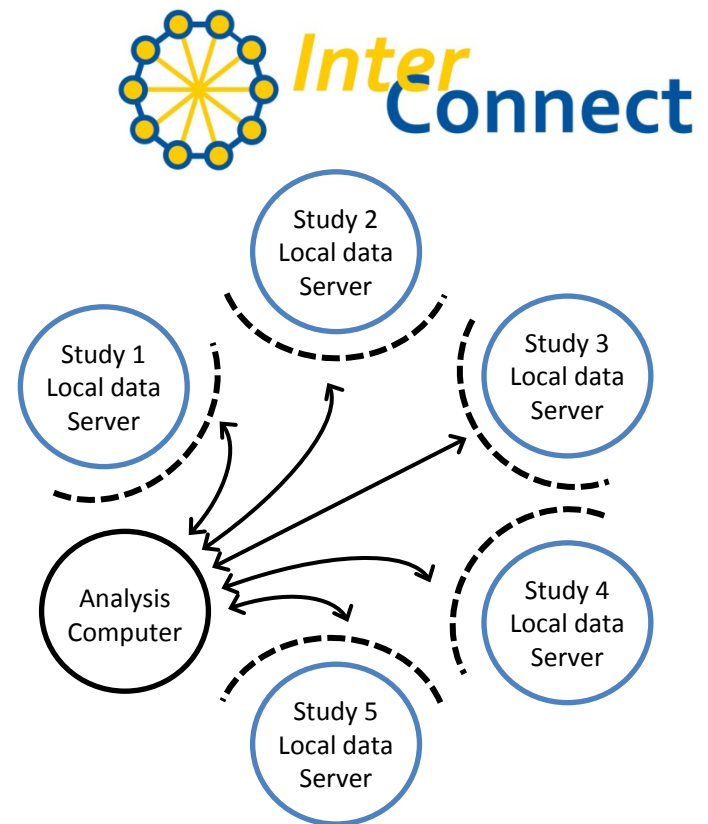
---

**Standard pooled analysis  
- base line**



**VS.**

**Bring analysis to the data  
- proof of concept**



# Do short-term vitamin D supplements prevent diabetes?

---



## Cambridge (n=172)

Randomised group encoded 0, 1, 2

HbA1c at baseline (%)

HbA1c at 4 months (%)

## London (n=168)

Randomised group encoded "Placebo", "Vit D2", "Vit D3"

HbA1c at baseline (mmol/mol)

HbA1c at 4 months (mmol/mol)

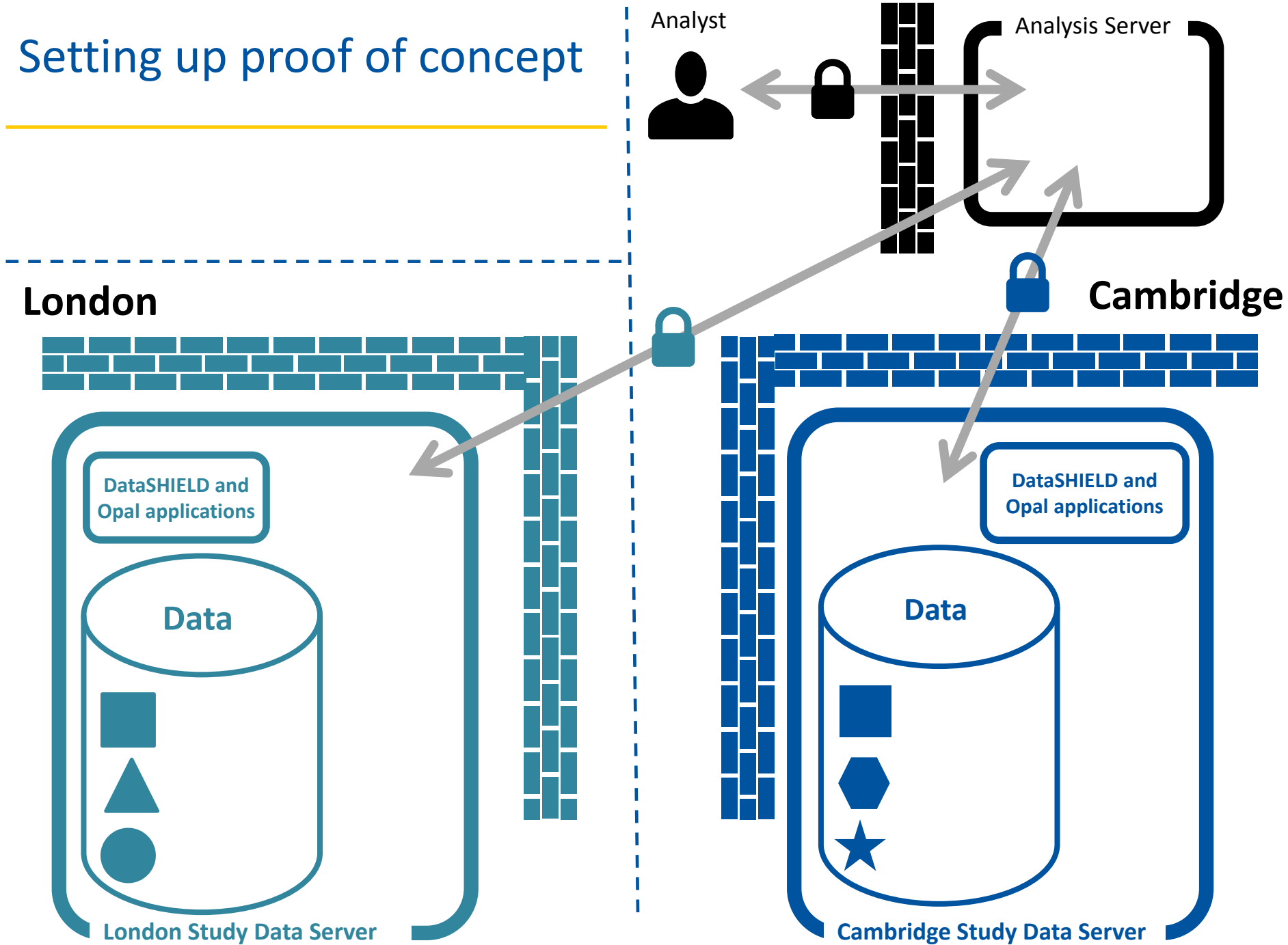
**Data made available thanks to:**

Stephen Sharp

Nita Forouhi

Graham Hitman

# Setting up proof of concept



# What is needed to set up a local data server and join InterConnect?

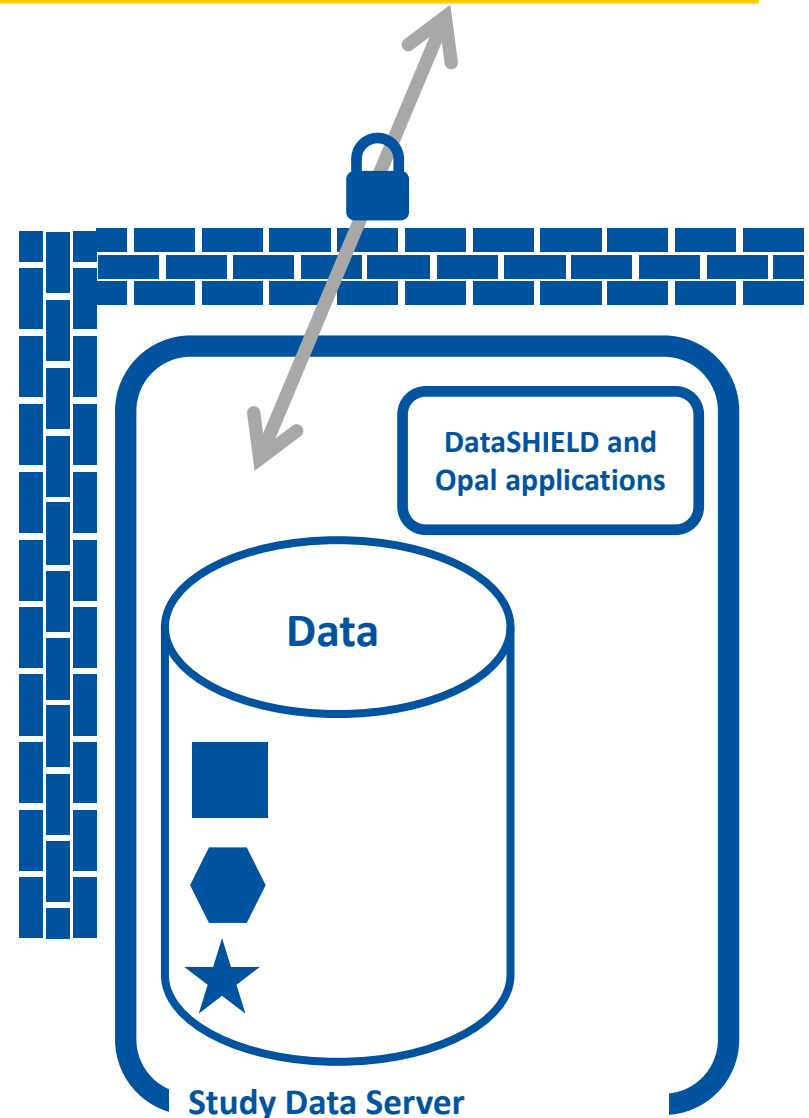
With support from InterConnect

The initial set up tasks **1** - **3** consist of standard work that could be managed by IT staff

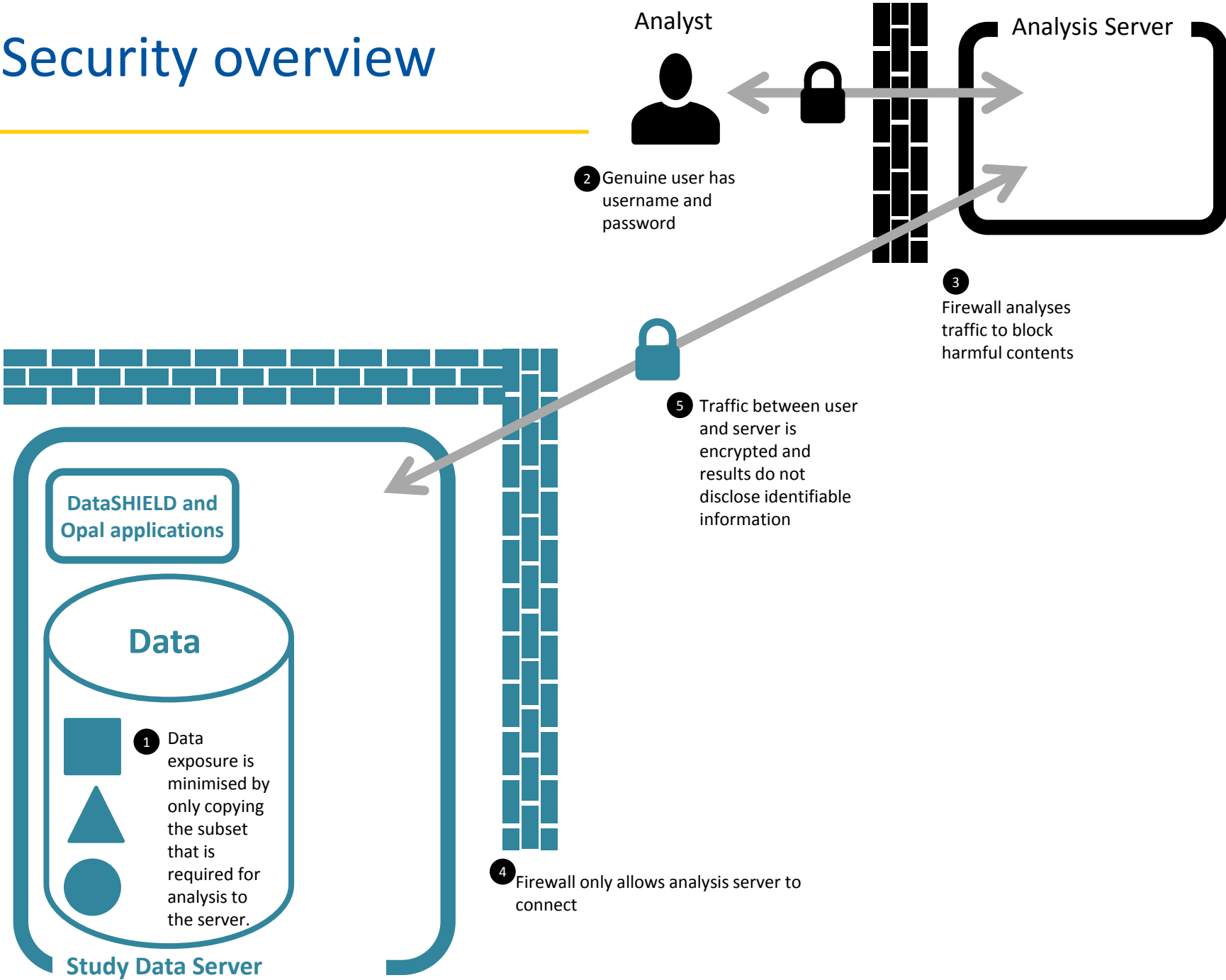
- 1** Obtain server hardware (could reuse existing system, use virtual machine or purchase new system)
- 2** Install operating system and configure basic settings
- 3** Configure institution's firewall to permit specific traffic to access the server

Tasks **4** - **7** could be managed by a researcher with some IT skills

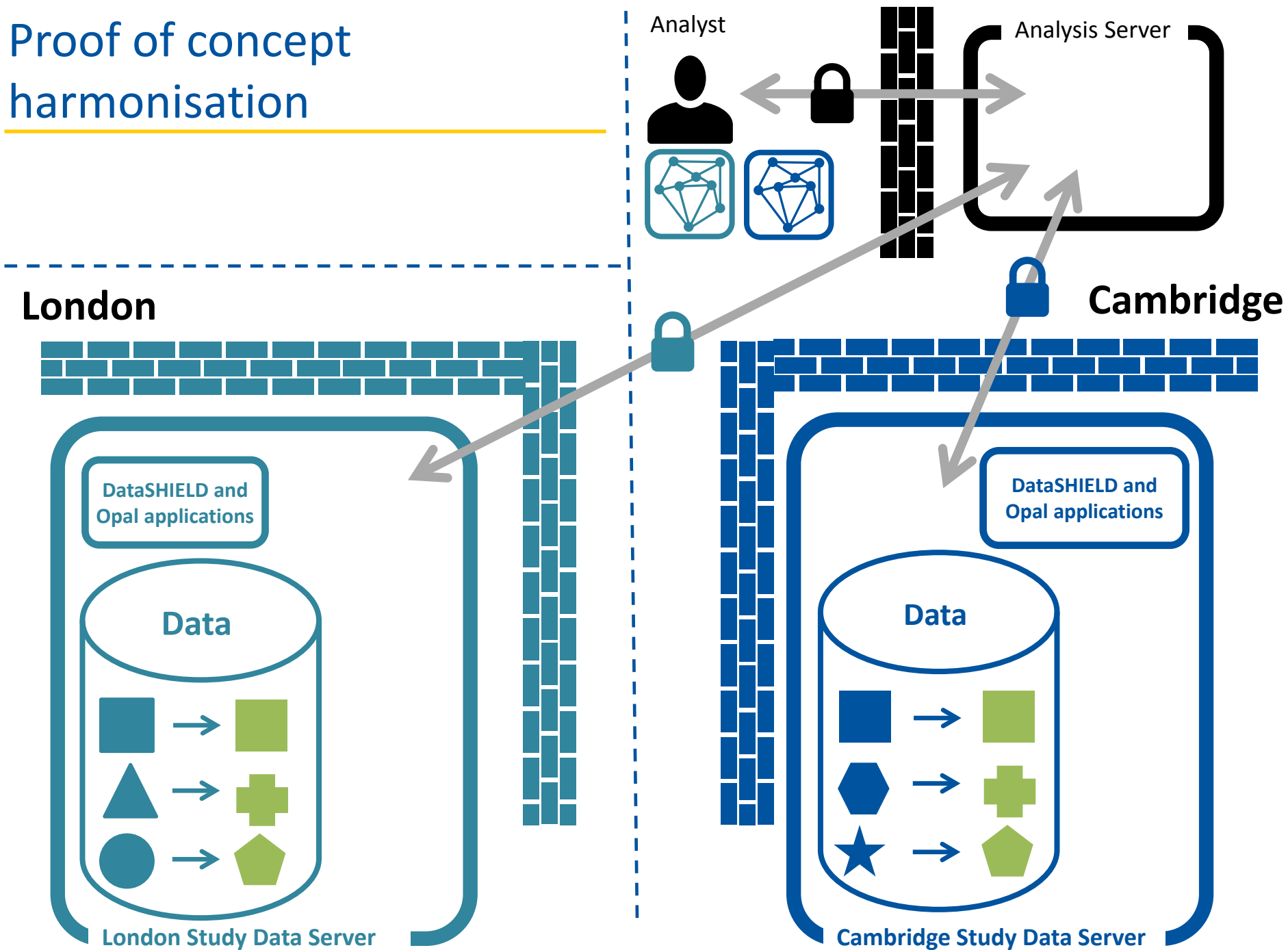
- 4** Install and configure data analysis software
- 5** Load relevant study data into database
- 6** Verify system by running tests
- 7** Manage updates for the software to ensure it stays up to date



# Security overview



# Proof of concept harmonisation





# Harmonisation process for one simple variable

---

1. Identify variables that require harmonisation

## Variable - HbA1c at 4 months

Cambridge - (%)

London - (mmol/mol)

2. Design the algorithm to align the variables

*“Multiply the value in mmol/mol by 0.09148 and add 2.152. If the value is missing, use 999.”*

3. Code algorithm in JavaScript on data server & capture in registry

```
var HBA1C_FU = $('hba1c4').multiply(0.09148).plus(2.152);  
if(HBA1C_FU.isNull().value()){  
  newValue(999, 'integer');  
}else{  
  HBA1C_FU;  
}
```

# Proof of concept analysis

London



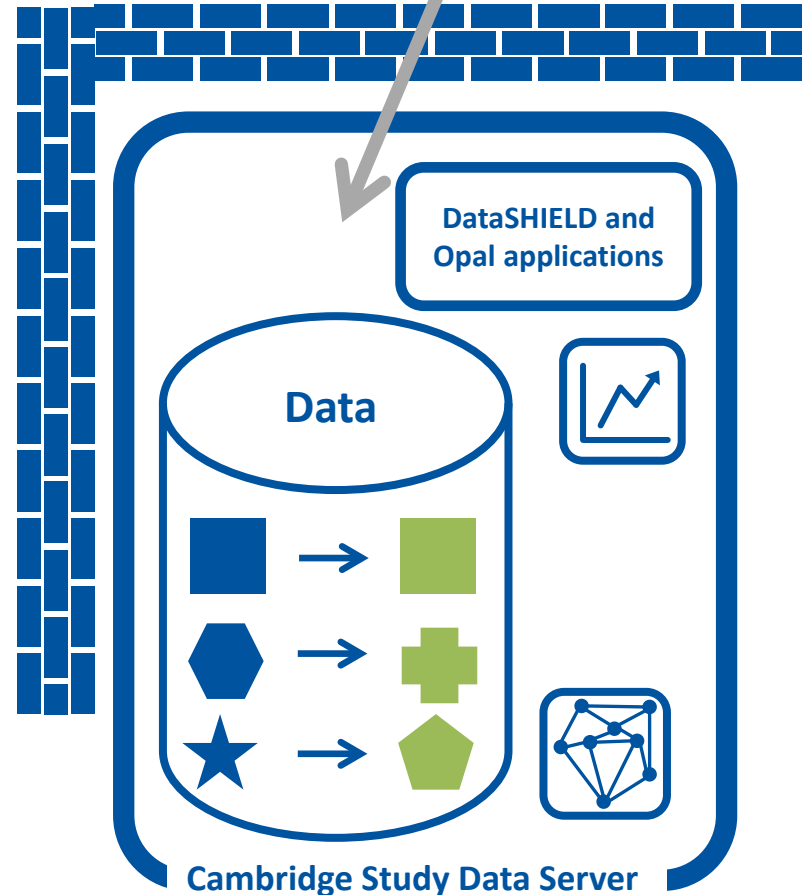
Analyst



Analysis Server



Cambridge



# Results from federated analysis

---

- The original pooled analysis showed no significant change in HbA1c when using vitamin D supplements and therefore don't prevent diabetes
- The federated analysis gave the same results as pooled analysis to 3 decimal places:

		<b>HBA1c %</b>	<b>low95CI</b>	<b>high95CI</b>	<b>p</b>
<b>Pooled analysis</b>	D2 vs placebo	-0.045	-0.104	0.015	0.14
	D3 vs placebo	0.018	-0.041	0.078	0.55
<b>Federated analysis</b>	D2 vs placebo	-0.045	-0.104	0.015	0.14
	D3 vs placebo	0.018	-0.041	0.077	0.55

# From proof of concept to exemplar

---

- Needed to develop new analysis functionality: Bespoke function developed successfully for pilot
- Challenges for lay user: Harmonisation algorithms in JavaScript, use of R and DataSHIELD
- Research exemplars will allow further development and knowledge transfer



**Inter**  
**Connect**



*Global data for diabetes and obesity research*

# Developing the vision via exemplar research questions

*Ken Ong*

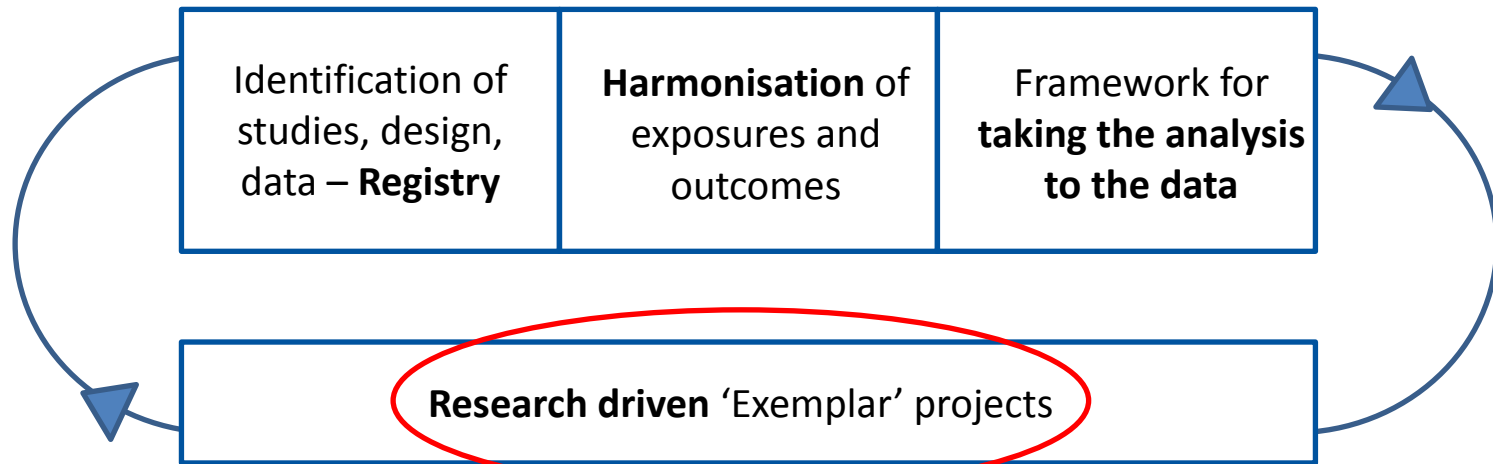
*InterConnect WP3 Leader & MRC Epidemiology Unit, University of Cambridge, UK*

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

# Implementation to drive development

---

## TOOLS & INFRASTRUCTURE



## RESEARCH USE: APPLICATION TO FOCUS & REFINE

*Forming ad hoc consortium – PA in pregnancy*

# Exemplar research question

---

“Is higher mother’s physical activity during pregnancy associated with lower offspring adiposity at birth?”

## Why is this important?

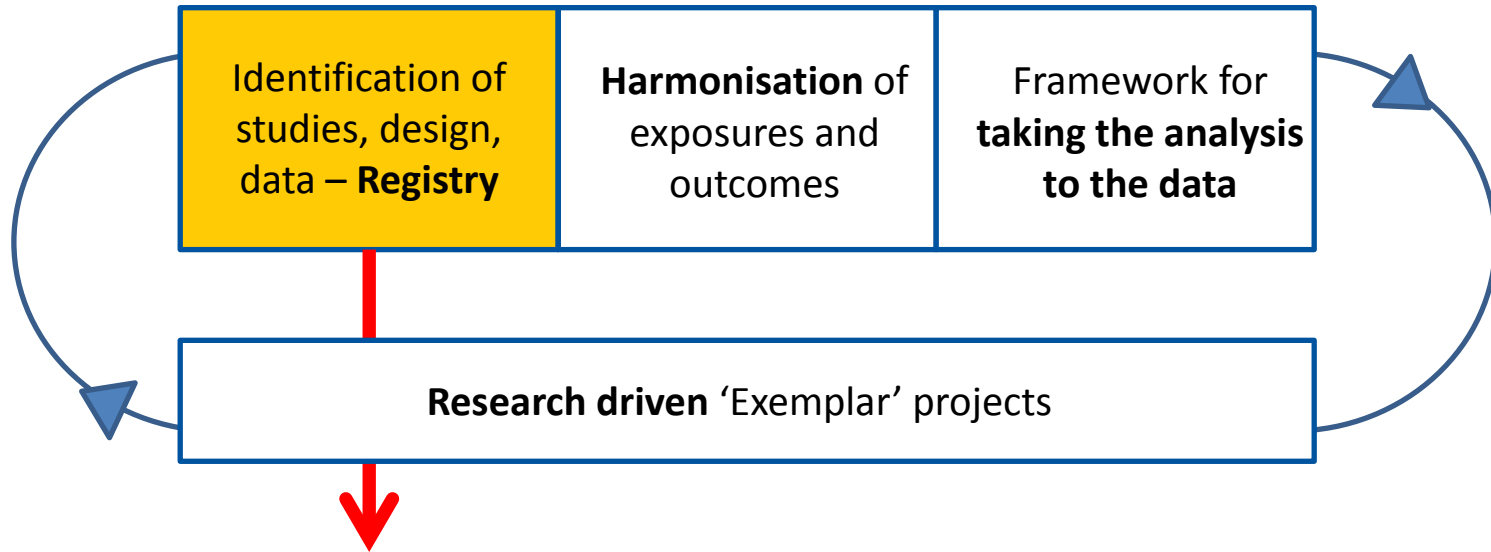
- Short-term risks of large baby for the mother & newborn
- Hypothesised long-term programming of metabolism in the offspring

# Existing evidence

---

- Variable impact of mother's physical activity on birth weight; and limited evidence on newborn adiposity
- Suggested greater impact in overweight and obese mothers, who have higher risks of large babies
- Impact also suggested to differ by modality (weight-bearing) and offspring sex





## We identified relevant studies by

- Contacting known investigators
- Searching review articles, own literature searches
- (in future – search the Registry)

# Forming an *ad hoc* consortium

---

- Discussed with a number of cohorts
  - Interest and intention to participate
- Held a Webex meeting
  - Explain the InterConnect vision
  - Collectively discussed practical issues and addressed FAQs

# Frequently asked questions

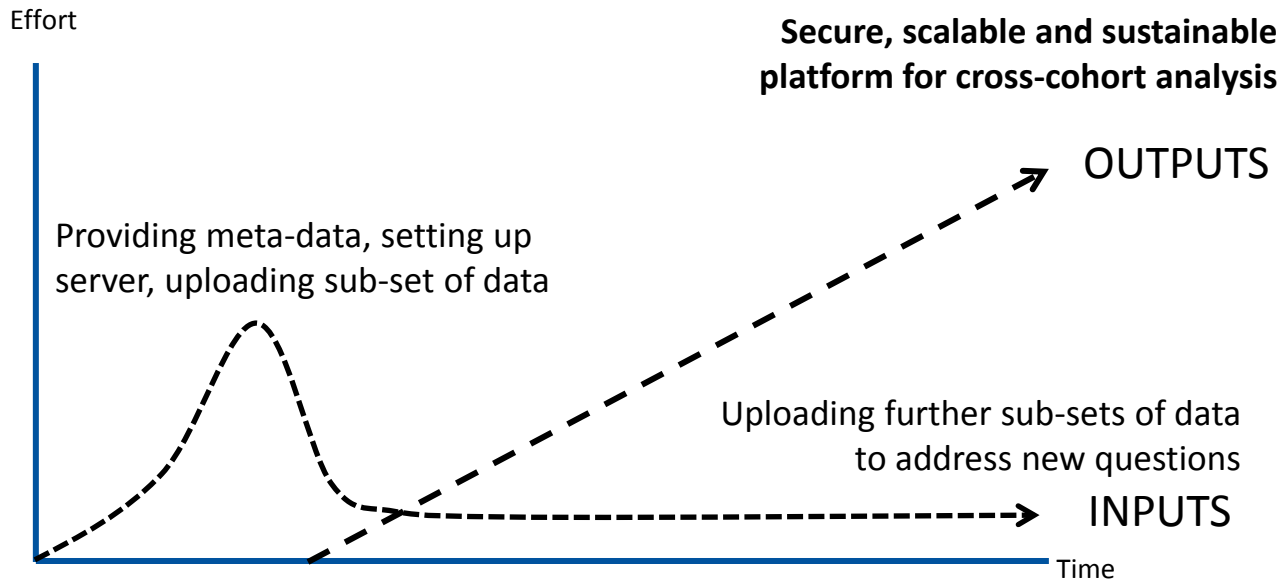
---

- IT set up and data security?
- Is it worth the upfront investment?
- Will I lose control of my data?
- What are the ELSI considerations?
- What is the publication policy?
- What is involved? Who does what?

# Is it worth the up-front investment?

---

- Once set up, re-use for further research questions
- Consortium is forming around first exemplar question
- Will then define further questions itself

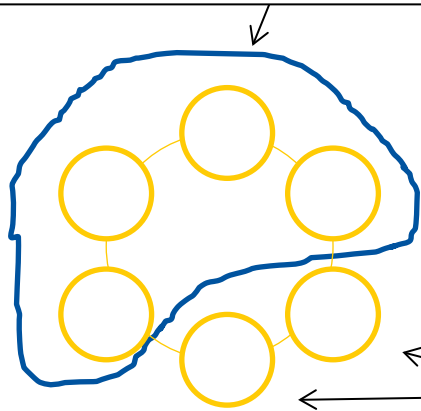


# Will I lose control of my data?

---

- No – the data is behind your local server firewall
- You control the access and the analyses undertaken

- Some studies agree to collaborate to address question A (consortium 1)
- IT permissions are set to allow remote access i.e. it is an active process
- This makes the relevant sub-set of data accessible
- Permissions can be revoked by the institution owning the data



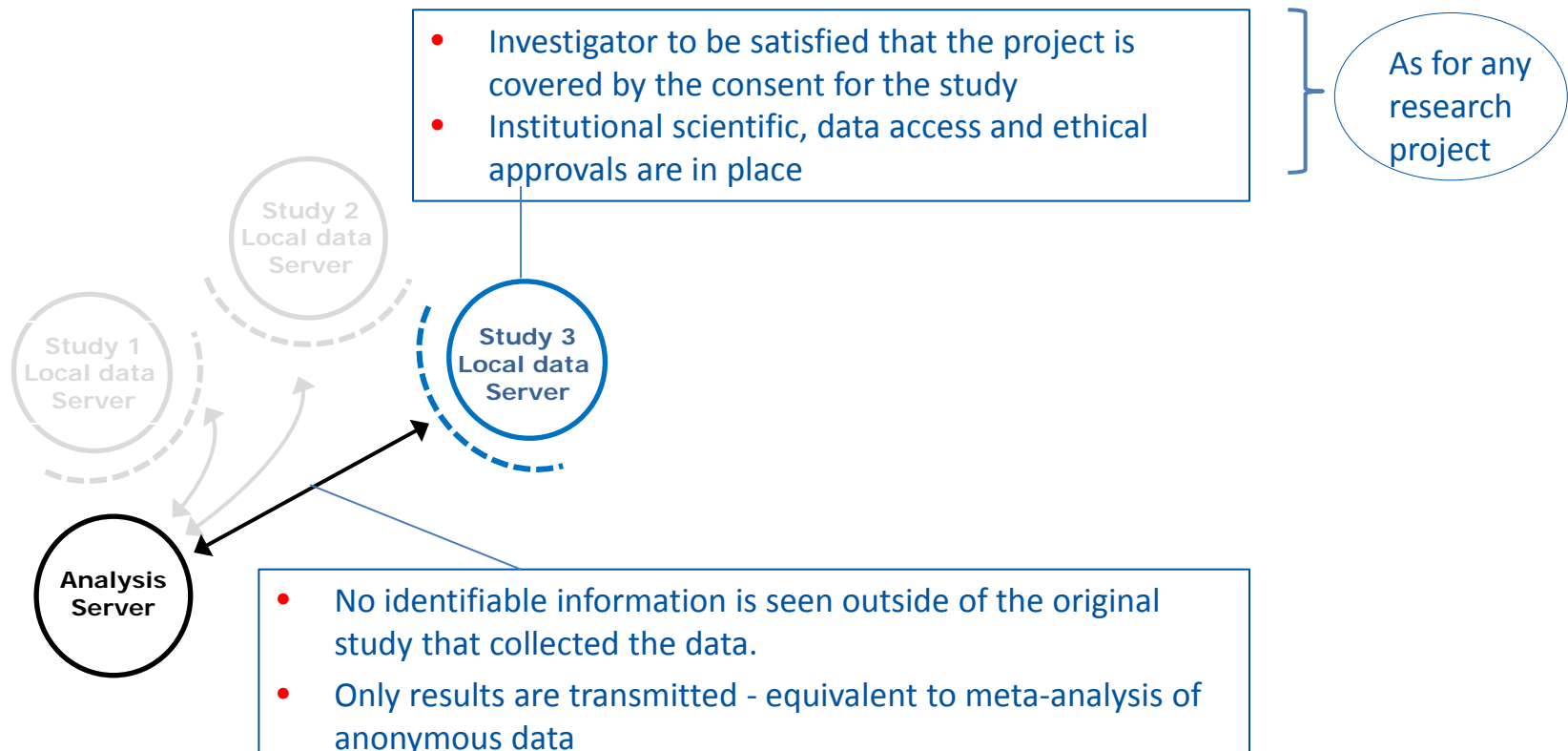
Consortium A  
Question 1

- Other studies do not wish to participate
- No IT permissions are put in place
- Their data are not available for this analysis

# What are the ELSI considerations?

---

- The data does not leave the institution
- As with any research, the study investigators are responsible for local approvals for the research question



# What is the publication policy?

---

- The publication policy is for each *ad hoc* consortium to decide

# What's involved, who does what?

---

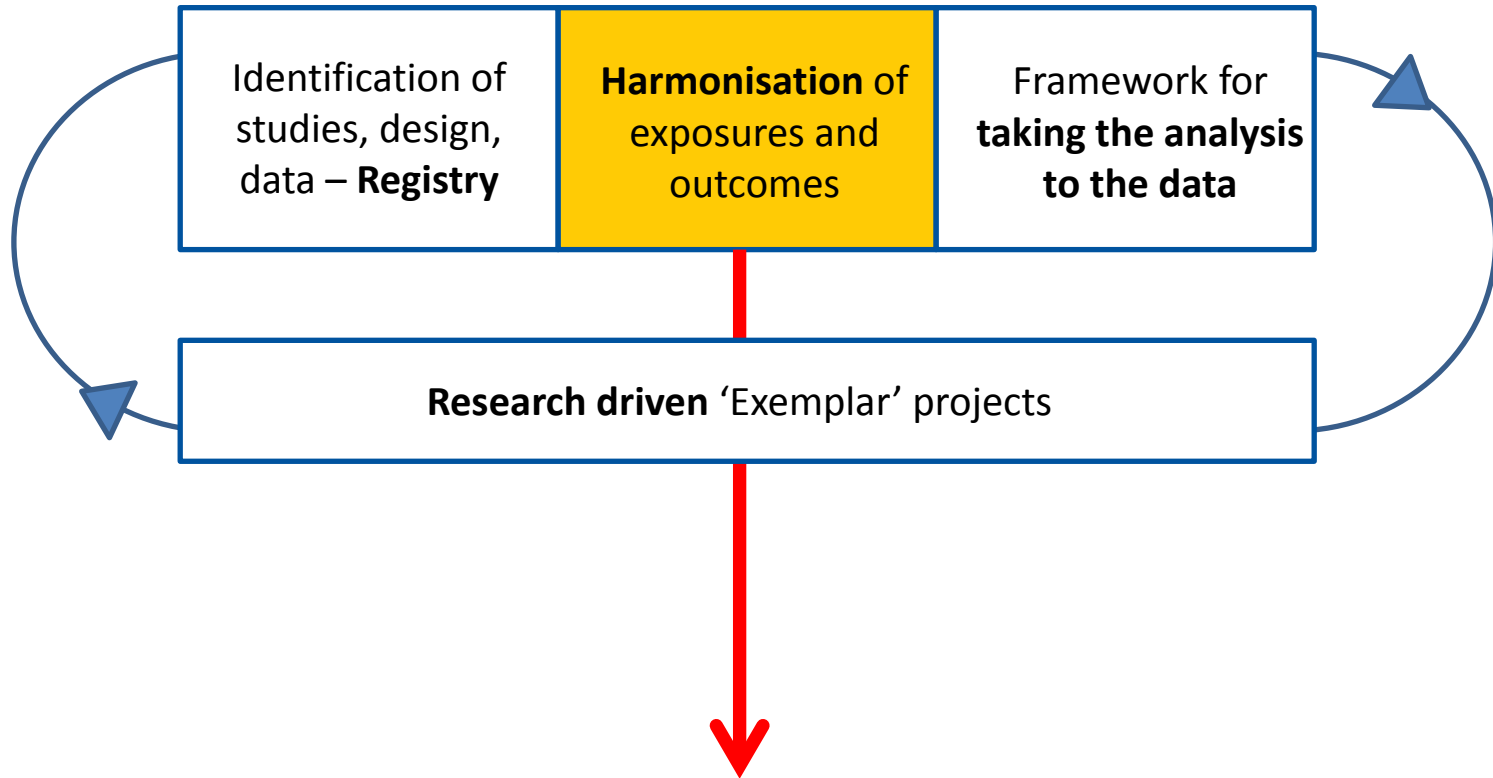
	Study Team	InterConnect Team (role)
Provide meta-data	✓	
Set up local server	✓	(✓) (Tech. support)
Upload relevant data to local server	✓	(✓) (Tech. support )
Decide how to harmonise data	✓	✓ (Lead)
Develop harmonisation algorithms		✓ (Lead)
Analyse data remotely		✓ (Lead)

*Studies can take on these roles in due course*



Study name	N	Country	PA measure	When measured?
ALSPAC	14,541	UK	Questionnaire	18w and 34w
ABCD study	8,266	Netherlands	Questionnaire	15.6w
DNBC	101,042	Denmark	Computer-assisted telephone interview	12w and 30w
Healthy Start Study	2,820	USA	Questionnaire interview	17w, 27w, 1d post-delivery
ROLO	800	Ireland	Questionnaire	First antenatal visit
SWS	12,583	UK	Questionnaire interview	Pre-pregnancy, 11w and 34 w

PHYSICAL ACTIVITY QUESTIONS	ROLO	ALSPAC	ABCD	Healthy start study	DNBC	SWS
<b>LEISURE/EXERCISE ACTIVITIES</b>						
Strenuous exercise	Y					Y
Moderate exercise	Y	Y				Y
Mild exercise	Y					Y
Play any sport/exercise		Y	Y		Y	
Asked for specific sports/activities	Y		Y	Y	Y	
Frequency	Y	Y	Y	Y	Y	Y
Duration	Y	Y	Y	Y	Y	Y
<b>SEDENTARY ACTIVITIES</b>						
Sitting	Y			Y		Y
Watching TV/computer games	Y			Y	Y	Y
Sleeping/Lying						Y
<b>WORK</b>						
PA at work assessed	Y		Y	Y	Y	
Heavy lifts		Y		Y	Y	
Walking			Y	Y	Y	
Standing					Y	
sitting				Y	Y	
<b>HOUSEHOLD ACTIVITIES</b>						
Household activities assessed	Y			Y		
Heavy household activities	Y			Y		
Lift heavy objects		Y				
<b>TRAVEL</b>						
Travel mode assessed	Y	Y		Y		
Walking	Y	Y		Y		
Cycling	Y	Y				

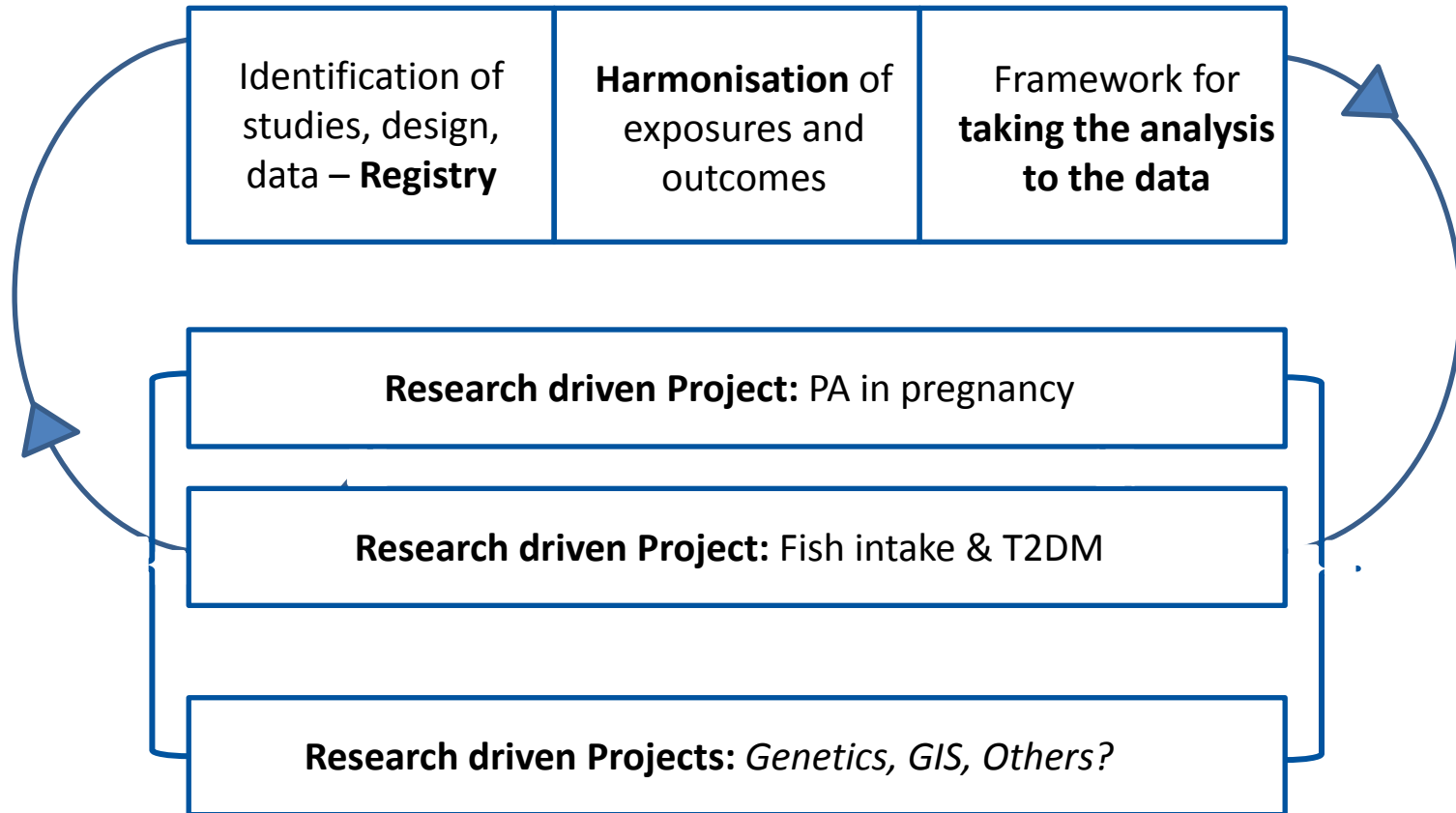


## Federated analysis allows flexible options

- Lowest common denominator approach (*e.g. collapse data to fit study with fewest categories of PA*)
- Estimate Latent Variables (*e.g. PA energy expenditure, intensity*)

# Future projects will drive future utility

---



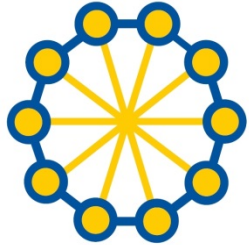


*Global data for diabetes and obesity research*

# Open discussion and involvement

*Nick Wareham*

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.



**Inter**  
**Connect**



*Global data for diabetes and obesity research*

## Acknowledgement

- This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

## Connect with us

- [InterConnect@mrc-epid.cam.ac.uk](mailto:InterConnect@mrc-epid.cam.ac.uk)
- [www.interconnect-diabetes.eu](http://www.interconnect-diabetes.eu)