Flexible disclosure control for microdata using DataSHIELD

Paul Burton

Data to Knowledge Research Group

University of Bristol, *School of Social and Community Medicine*

University of Newcastle, *Institute of Health & Society*

McGill University, OICR, Maelstrom Research

MRC Epidemiology Unit, Cambridge

CLOSER KEW – Disclosure Control: London, 18th January, 2017

Microdata = individual level data = individual patient data (IPD)

- Microdata are **absolutely fundamental** to contemporary science including biomedical, social, and public health science
- For someone wishing to analyse, interpret and draw conclusions from data they provide
  - The only way to do certain analyses
  - Enhanced efficiency in some circumstances
  - Greater flexibility

Constraints and barriers to sharing and combining microdata

- Ethico-legal or other governance restrictions
- Maintaining control of intellectual property
- Physical size of data

The DataSHIELD approach

- Take “analysis to data” ….. not “data to analysis”
- Leave the data to be analysed on local servers behind the firewalls where they usually reside
- The analysis centre *co-ordinates* parallelised analyses in all studies simultaneously
- Tie analysis chains together with non-disclosive information
- Analytic processing - *and options for disclosure control* - located with the data

DataSHIELD:
Data Aggregation Through Anonymous Summary-statistics from Harmonized Individual-levEL Databases

- **Horizontal partitioning**
  - **Different sources hold all variables but on different individuals**
  - **Secure meta-analysis (IPD and Study-Level)**
  - **Secure single-site analysis**
- **Vertical partitioning**
  - **Different sources hold different variables on the same individuals**
  - **Secure processing and analysis of linked data without bringing the data together**

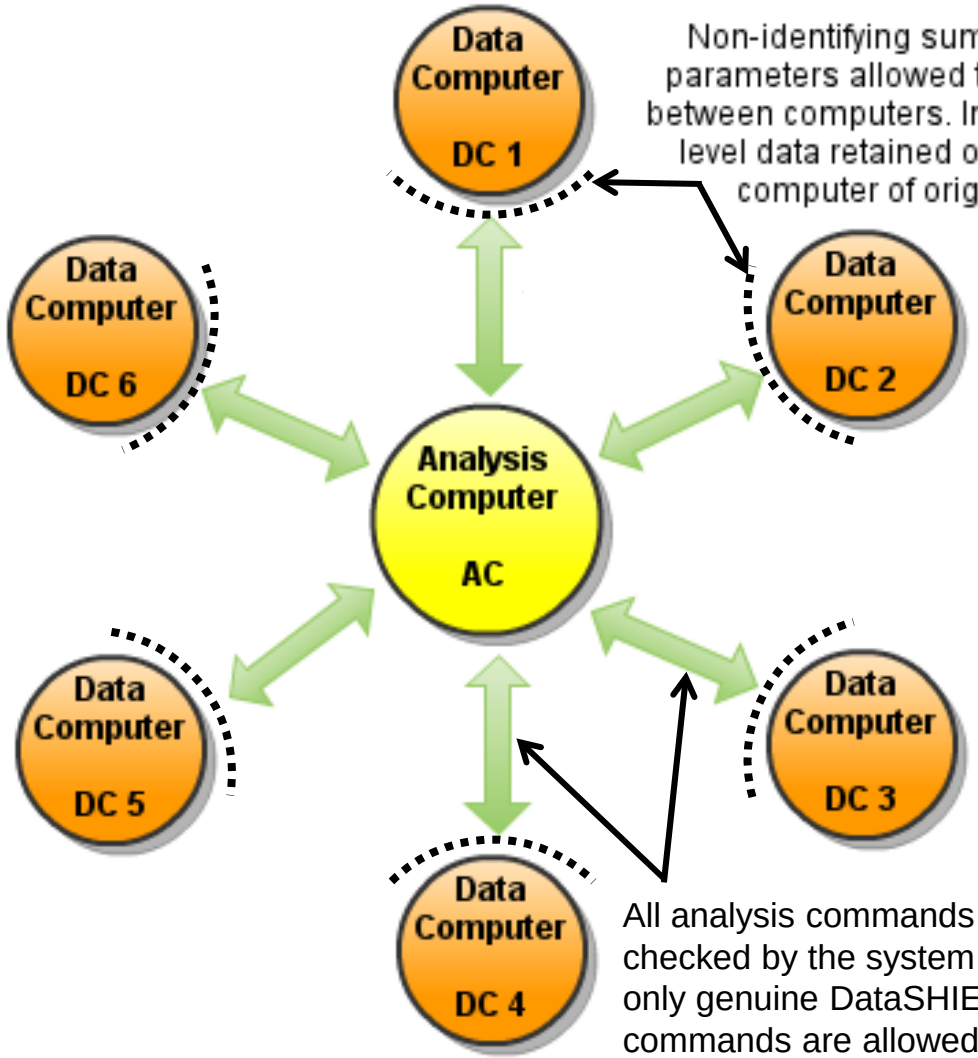
D2K data to knowledge

SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration

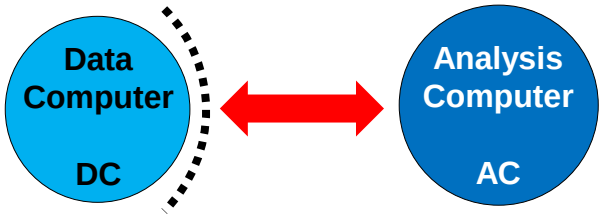
2009: The DataSHIELD challenge

Given that microdata are scientifically critical and yet potentially sensitive, can we ensure that the information driving analysis only ever emerges from the firewall of each data source in non-disclosive form? (i) encryption (trivial and non-trivial); (ii) low dimensional (ideally sufficient) statistics

Multi-site DataSHIELD horizontally partitioned data



Single site DataSHIELD

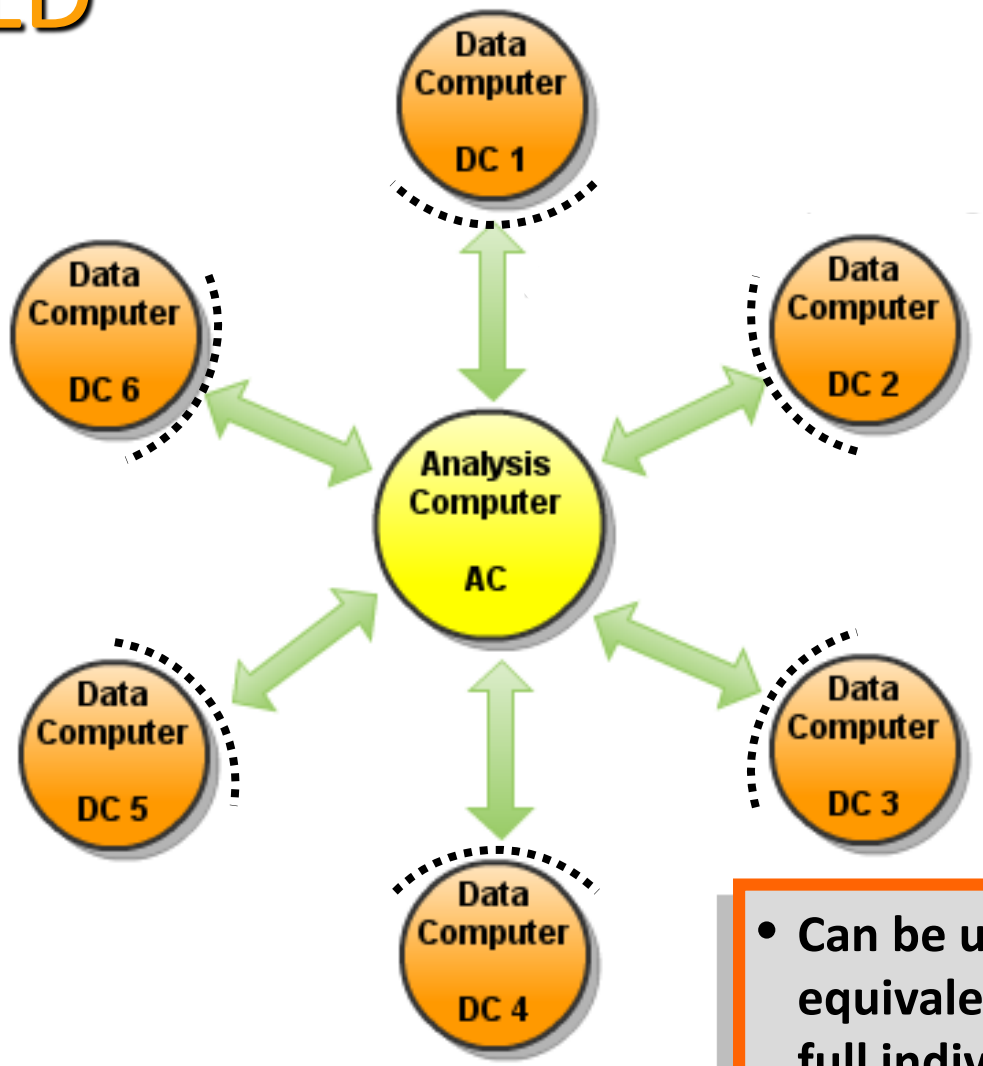


SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration

D2K data to knowledge

The DataSHIELD solution

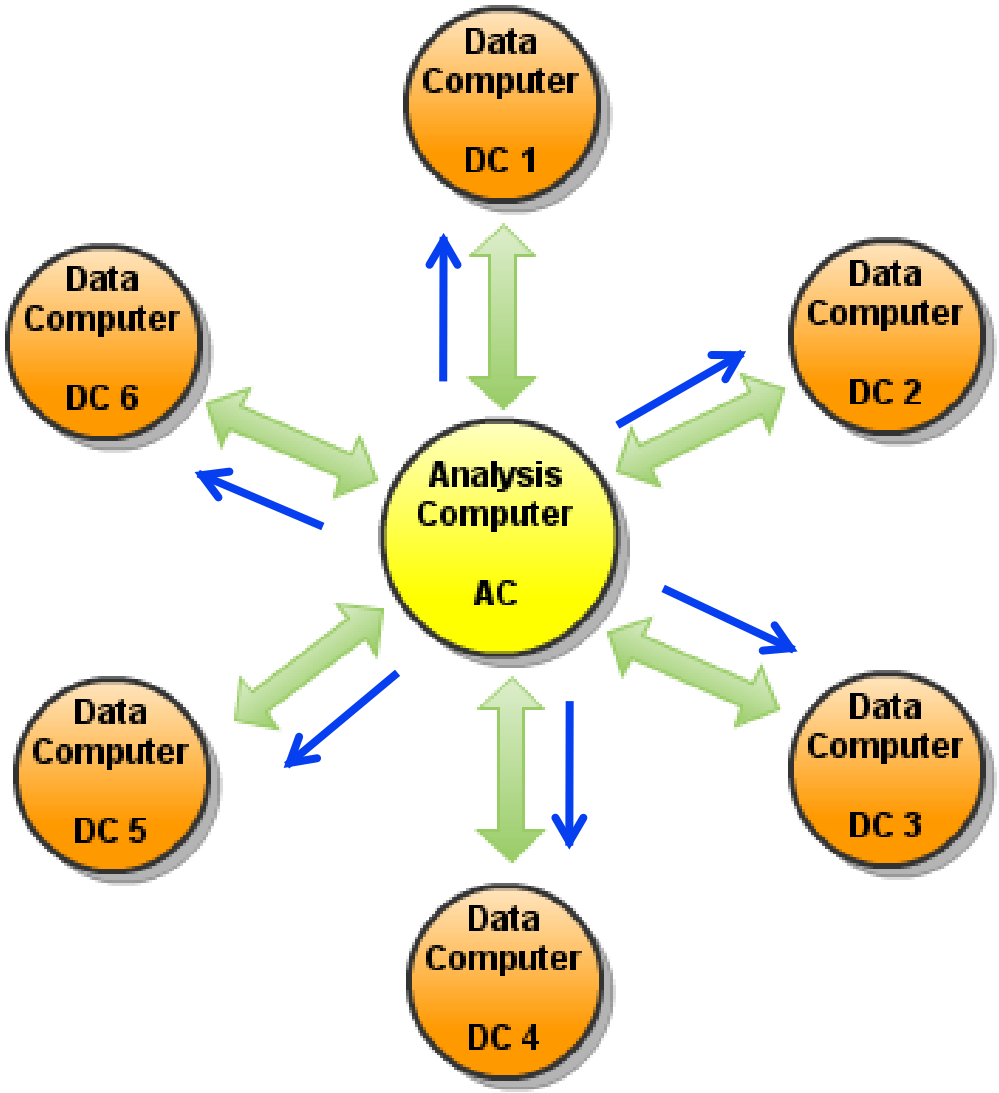
- One step analyses: *e.g.* ds.table2D - request non-disclosive output from all sources
- Multi-step analyses: *e.g.* ds.lexis – set up and then request output
- Iterative analyses: *e.g.* ds.glm - parallel processes linked together by non-identifying summary statistics – e.g. for glm = score vectors and information matrices



- Can be used as equivalent to full individual level analysis or to study level meta-analysis

SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration D2K data to knowledge

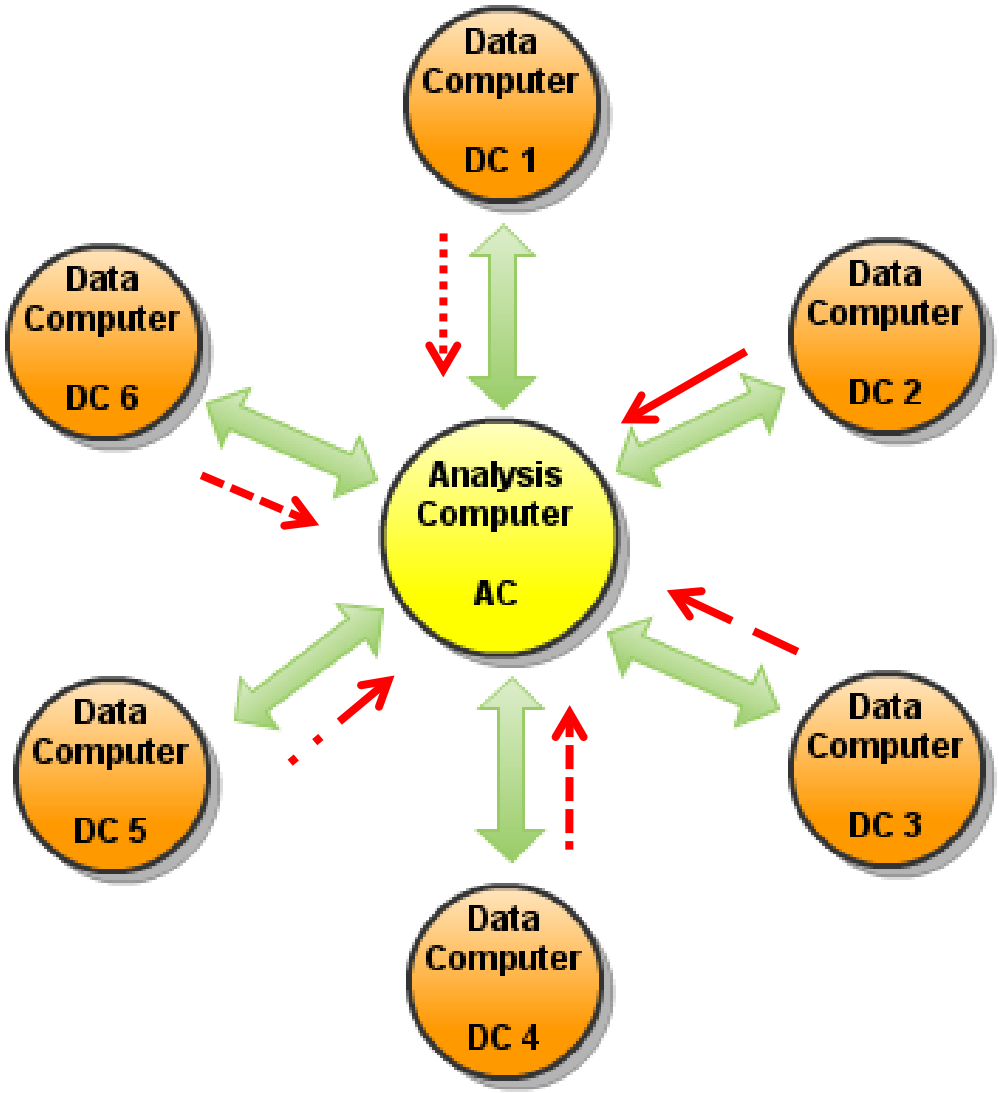
DataSHIELD



Analysis commands (1)

```
b.vector<-c(0,0,0,0)

glm(cc~1+BMI+BMI.456+SNP,
family=binomial,
start=b.vector, maxit=1)
```

DataSHIELD



Summary Statistics (1)

Score vector Study 5

[36, 487.2951, 487.2951, 149]

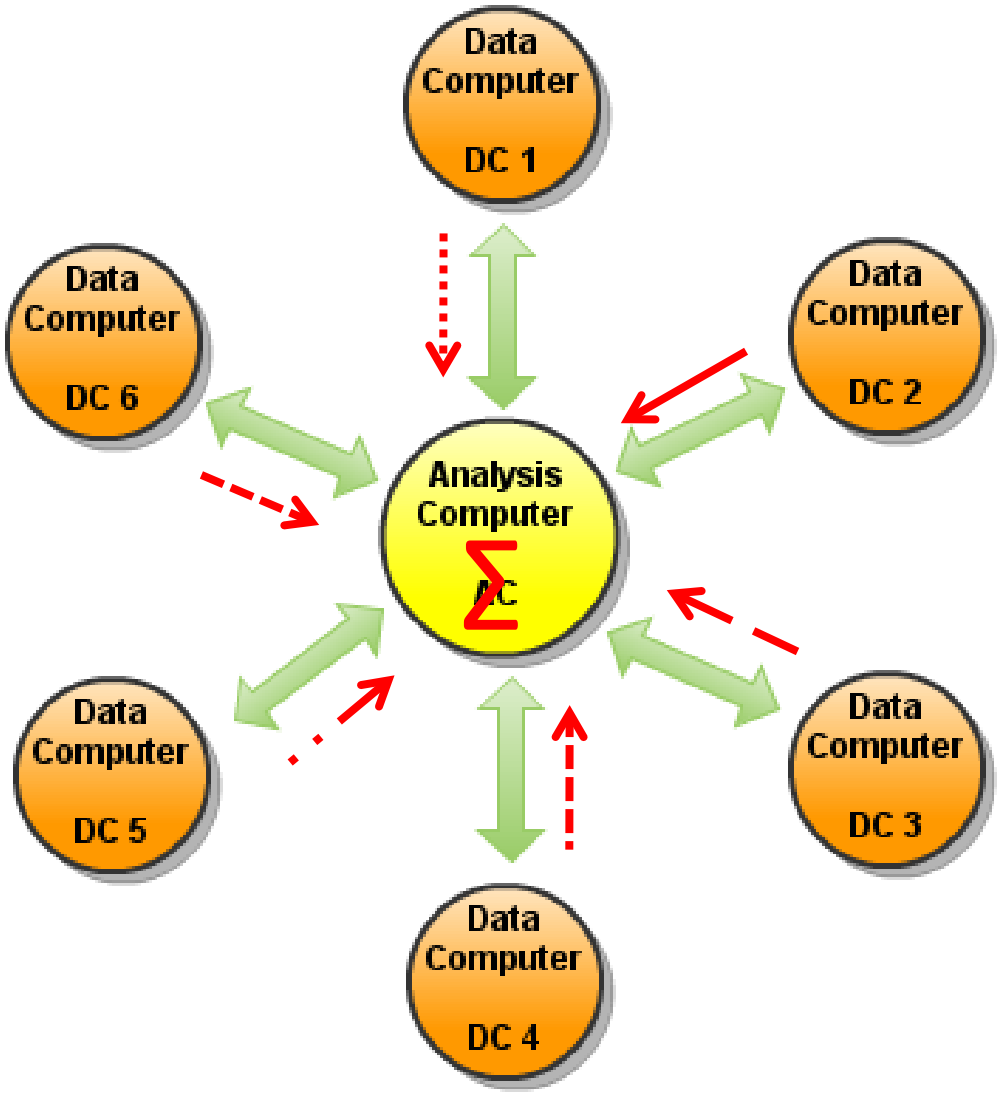
Information Matrix Study 5

500	70.56657	70.56657	297
70.56657	7646.29164	7646.29164	65.39412
70.56657	7646.29164	7646.29164	65.39412
297	65.39412	65.39412	382

D2K data to knowledge

SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration

DataSHIELD



Summary Statistics (1)

Score vector Study 5

[36, 487.2951, 487.2951, 149]

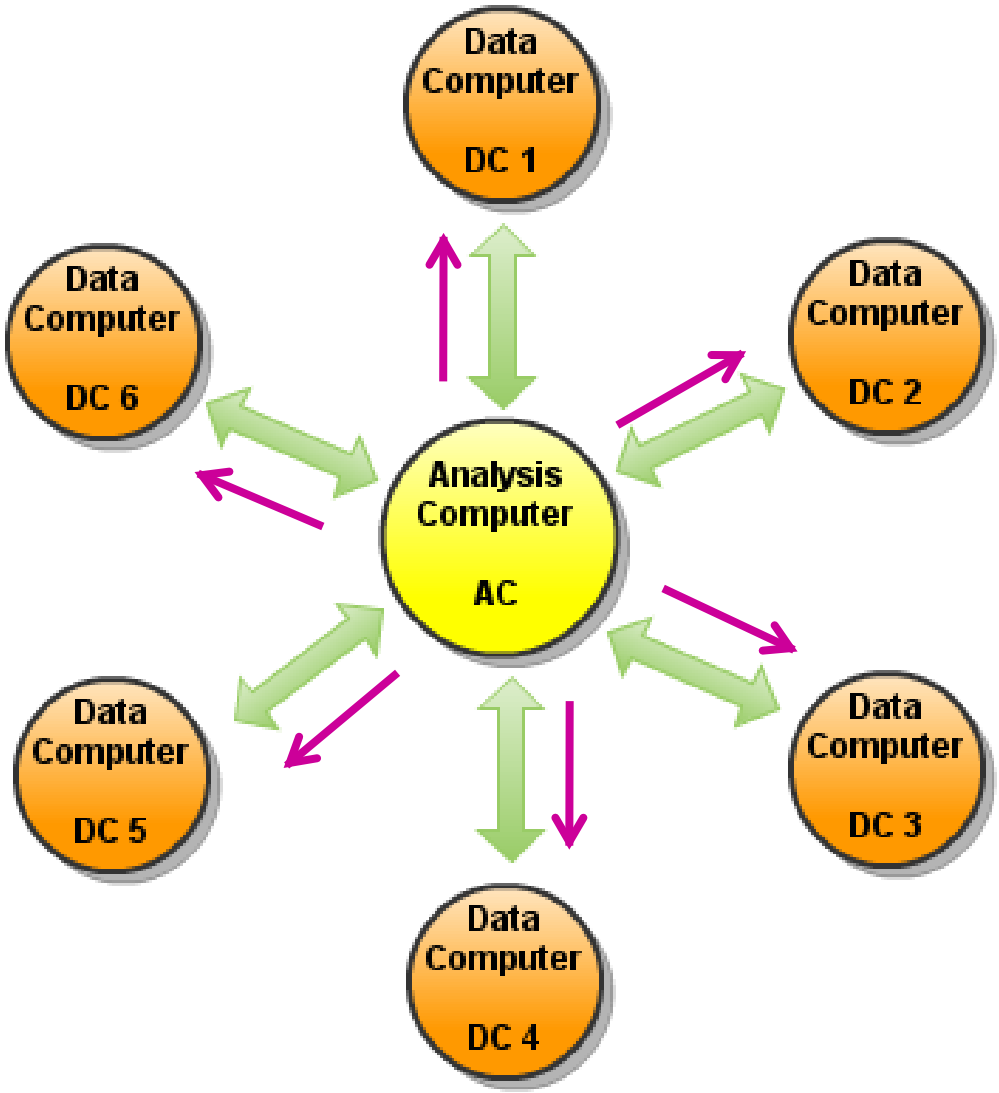
Information Matrix Study 5

500	70.56657	70.56657	297
70.56657	7646.29164	7646.29164	65.39412
70.56657	7646.29164	7646.29164	65.39412
297	65.39412	65.39412	382

D2K data to knowledge

SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration

DataSHIELD

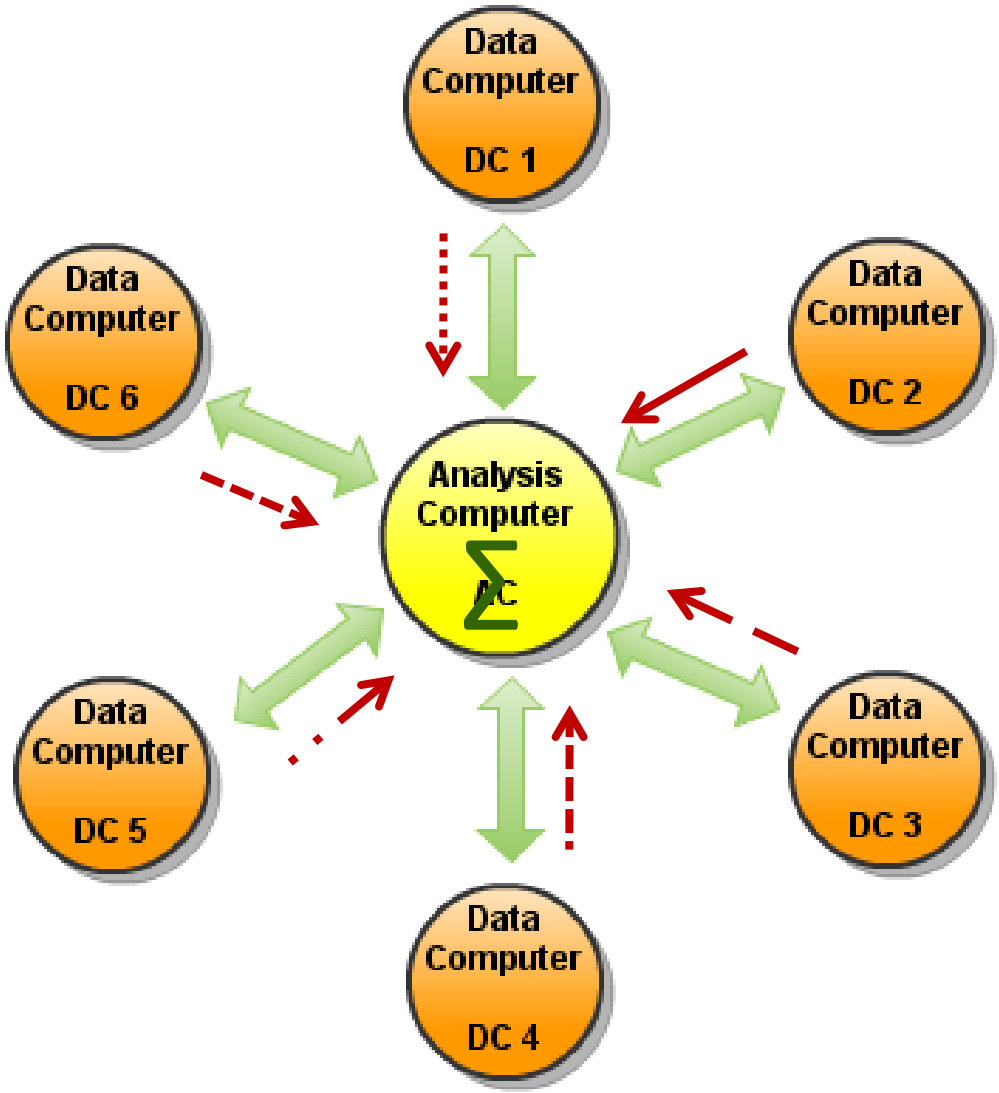


Analysis commands (2)

```
b.vector<-
c(-0.322, 0.0223, 0.0391, 0.535)

glm(cc~1+BMI+BMI.456+SNP,
family=binomial,
start=b.vector, maxit=1)
```

and so on .....

DataSHIELD



Updated parameters (4)

Final parameter estimates

Coefficient	Estimate	Std Error
Intercept	-0.3296	0.02838
BMI	0.02300	0.00621
BMI.456	0.04126	0.01140
SNP	0.5517	0.03295

D2K data to knowledge

SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration

Direct conventional analysis

Coefficients:

	Estimate	Std. Error
(Intercept)	-0.32956	0.02838
BMI	0.02300	0.00621
BMI.456	0.04126	0.01140
SNP	0.55173	0.03295

Does it work?

DataSHIELD analysis

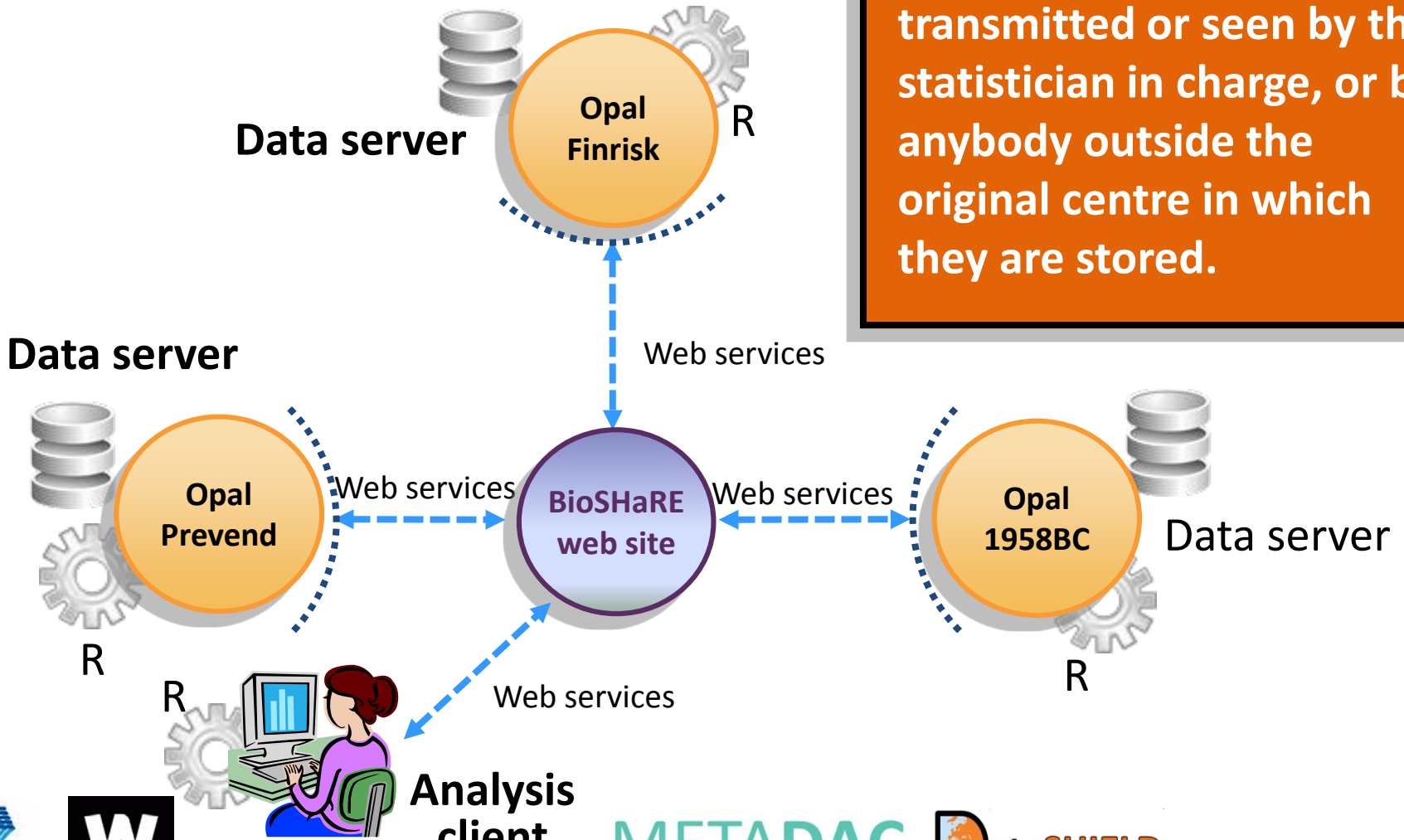
Parameter	Coefficient	Standard Error
$b_{intercept}$	-0.3296	0.02838
b_{BMI}	0.02300	0.00621
$b_{BMI.456}$	0.04126	0.01140
b_{SNP}	0.5517	0.03295

DataSHIELD: current implementation for horizontally partitioned data

Server-side functions

Client-side functions

Individual level data never transmitted or seen by the statistician in charge, or by anybody outside the original centre in which they are stored.



SEVENTH FRAMEWORK

wellcome

METADAC

DataSHIELD Secure Bioscience Collaboration

D2K data to knowledge

Current research activity

- InterConnect
- SPIRIT
- ENPADASI
- BioSHaRE-EU Healthy Obese Project
- BioSHaRE-EU Environmental Core Project

- Farr Institutes
- UKDA
- F1000 – Research Journal

D2K data to knowledge

SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration

Important reflections

- DataSHIELD is aimed at allowing governance constraints to be met in full while the governance bar is (rationally) lowered? This can streamline data access and promote open science.
- DataSHIELD provides an effective solution to a range of challenges in data management. It is not always appropriate and may well be associated with unnecessary costs, workload and time delay if it is used in inappropriate circumstances. Inappropriate settings may include situations where DataSHIELD's disclosure protection is too weak or unnecessarily rigorous.
- DataSHIELD should ideally be applied on top of a data access and governance system that is already well founded and a hardware/middleware infrastructure that is already robust and resilient *e.g.* resistant to hacking
- It is recommended that the Opal servers that contain the data to be analysed using DataSHIELD should be kept separate from the servers holding the main data systems for a study and data should be pseudonomized

Important reflections

- Long term sustainability demands the development and application of a cost-recovery mechanism to support support for implementation and use of Opal and DataSHIELD and for undertaking tailored development of new functionality to address the needs of particular projects
- Because DataSHIELD allows efficient access to the full information held in ethicolegally or intellectually sensitive microdata while those microdata physically remain with their original generator or formal custodian. It can therefore promote trust-based collaboration particularly in lower/middle income countries where there are understandable sensitivities about rich research groups from elsewhere “going off” with precious data and gaining most of the scientific return that those data may offer.

D2K data to knowledge

SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration

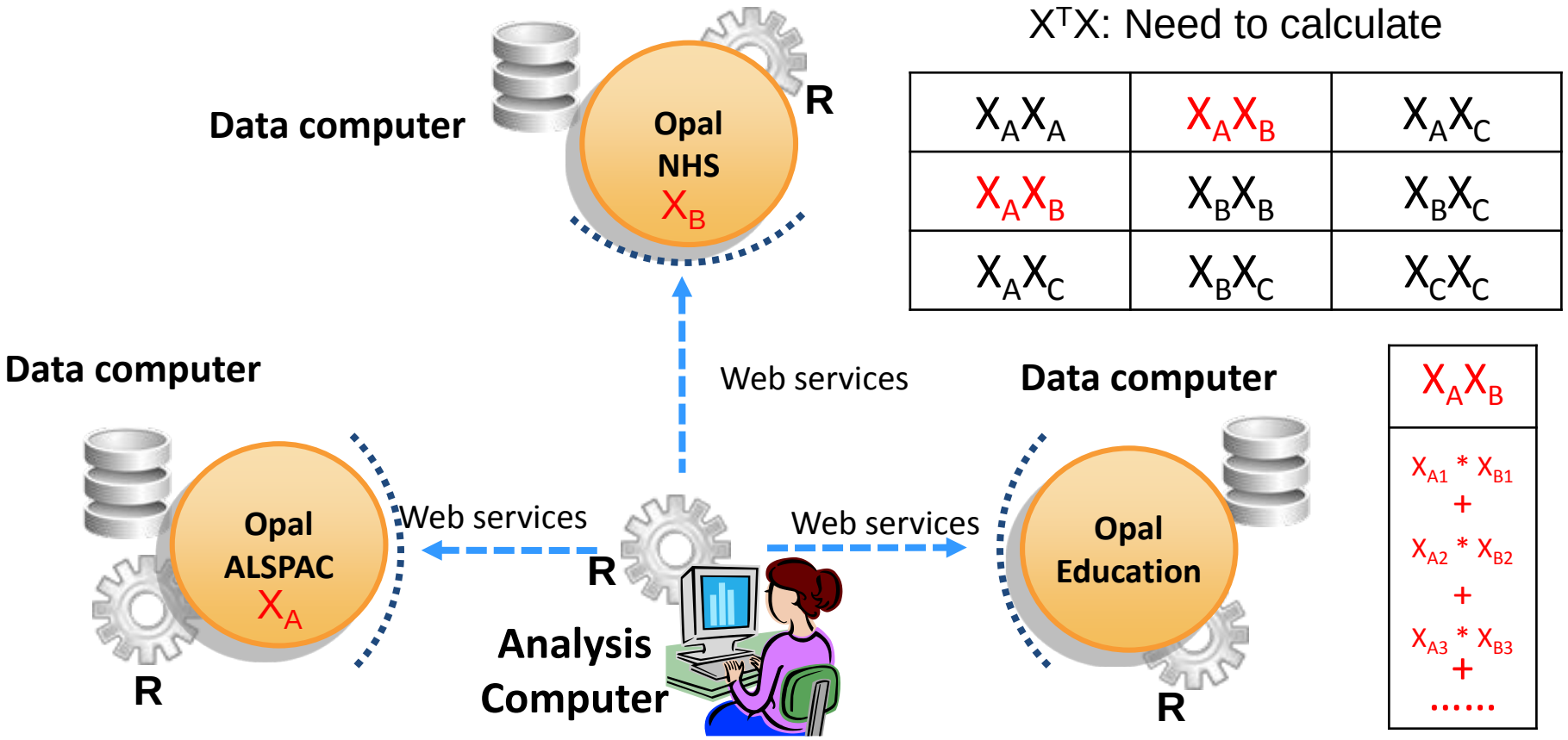
Important reflections

- Like any other approach to analysis – or joint co-analysis – there is little point in applying DataSHIELD to data unless they have first been cleaned and harmonized. If this is not done, results may actually be misleading. Initial data preparation can take much more work than the DataSHIELD analysis itself.
- Any rational overall data management strategy for biomedical and health data should take advantage of the complementary strengths of central warehousing and remote federated analysis. DataSHIELD is of particular value for the latter but also supports the former.
- Long term sustainability demands the development and application of a cost-recovery mechanism to support support for implementation and use of Opal and DataSHIELD and for undertaking tailored development of new functionality to address the needs of particular projects.
- Inferential disclosure based on analytic results presents a challenge for all forms of data release. This is as true for DataSHIELD as for any other approach and needs active exploration.

THANK YOU FOR LISTENING

DataSHIELD: current implementation for vertically partitioned (linked) data

Regression coefficients = X^TY / X^TX



X^TX: Need to calculate

X_AX_A	X_AX_B	X_AX_C
X_AX_B	X_BX_B	X_BX_C
X_AX_C	X_BX_C	X_CX_C

X_AX_B
$X_{A1} * X_{B1}$ + $X_{A2} * X_{B2}$ + $X_{A3} * X_{B3}$ + ...

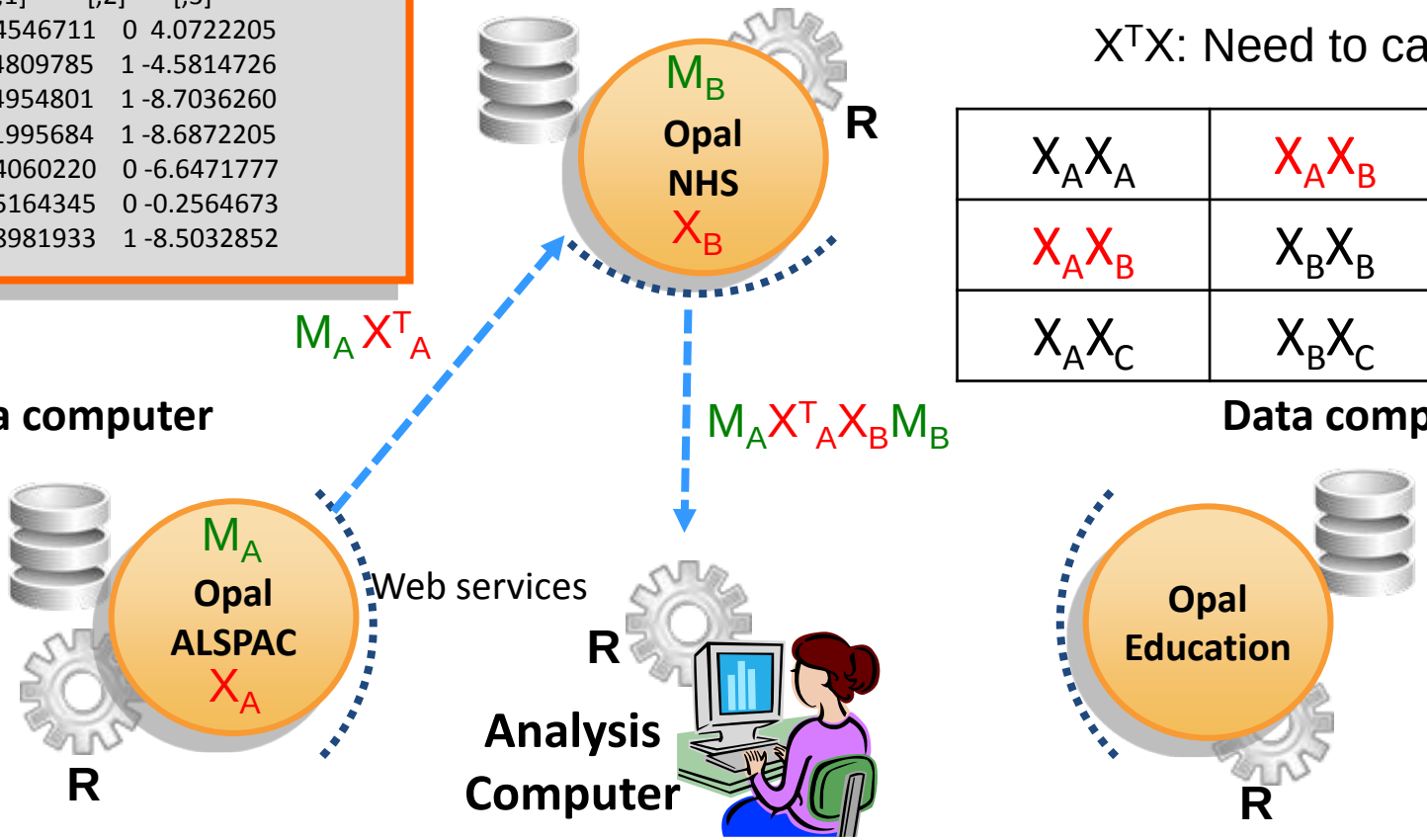
DataSHIELD: current implementation for vertically partitioned (linked) data

```
plain.text.vector.A  plain.text.vector.N
0 1 1 1 0 0 1        1 1 0 1 0 0 1

encryption.matrix
        [,1]       [,2]        [,3]
[1,] -1.444769  2.495677 -5.322736
[2,] -1.355529 -9.369041  2.687347
[3,]  4.603762 -3.622044 -2.817478

occluded.matrix.A
        [,1]   [,2]      [,3]
[1,] -1.4546711   0  4.0722205
[2,]  6.4809785   1 -4.5814726
[3,]  4.4954801   1 -8.7036260
[4,]  0.1995684   1 -8.6872205
[5,] -6.4060220   0 -6.6471777
[6,] -0.5164345   0 -0.2564673
[7,] -5.8981933   1 -8.5032852
```

Regression coefficients = X^TY / X^TX

X^TX: Need to calculate

X_AX_A	X_AX_B	X_AX_C
X_AX_B	X_BX_B	X_BX_C
X_AX_C	X_BX_C	X_CX_C



$$(M_A)^{-1} M_A X^T_A X_B M_B (M_B)^{-1} = X_A X_B$$

SEVENTH FRAMEWORK · wellcome · MRC Medical Research Council · E·S·R·C Economic & Social Research Council · METADAC · DataSHIELD Secure Bioscience Collaboration · D2K data to knowledge

Why DataSHIELD?

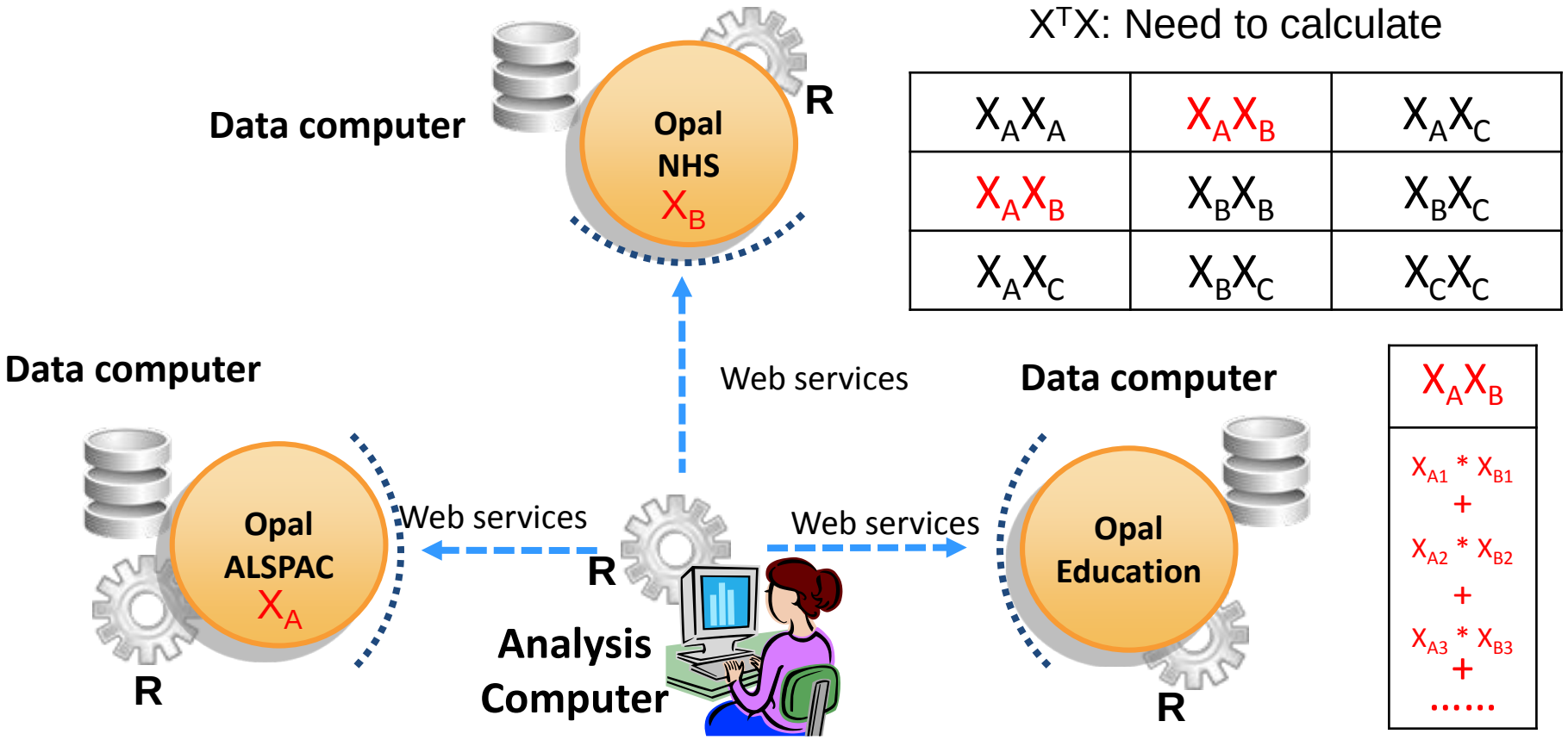
- **Horizontal DataSHIELD** multi-site
  - Secured IPD meta-analysis or study-level meta-analysis where different studies hold the same variables on different individuals
  - Data remains behind firewall of study holding data and is invisible and unobtainable externally. Appropriate disclosure settings under control of data controller. 'Local' study data storage could be at a national repository
  - Open-source freeware
- **Vertical DataSHIELD**
  - Ultra-secure analysis of very sensitive linked data where no source is prepared for its linked data to be held by any other sources or a trusted third party
  - Securing the linkage process itself
  - Non-linkage applications eg in 'omics (GCTA, kinship relationship matrix)
- **Single site Horizontal DataSHIELD (a freeware-based data enclave)**
  - Post-publication "open access" to sensitive data
  - "Open access" to simple descriptive stats from rigorously governanced studies
  - Analytic access to sensitive (but not ultra-sensitive) linked data
  - Analytic access to data collected by researchers in resource-poor regions

D2K data to knowledge

SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration

DataSHIELD: current implementation for vertically partitioned (linked) data

Regression coefficients = X^TY / X^TX



X^TX: Need to calculate

X_AX_A	X_AX_B	X_AX_C
X_AX_B	X_BX_B	X_BX_C
X_AX_C	X_BX_C	X_CX_C

X_AX_B
$X_{A1} * X_{B1}$ + $X_{A2} * X_{B2}$ + $X_{A3} * X_{B3}$ + ...

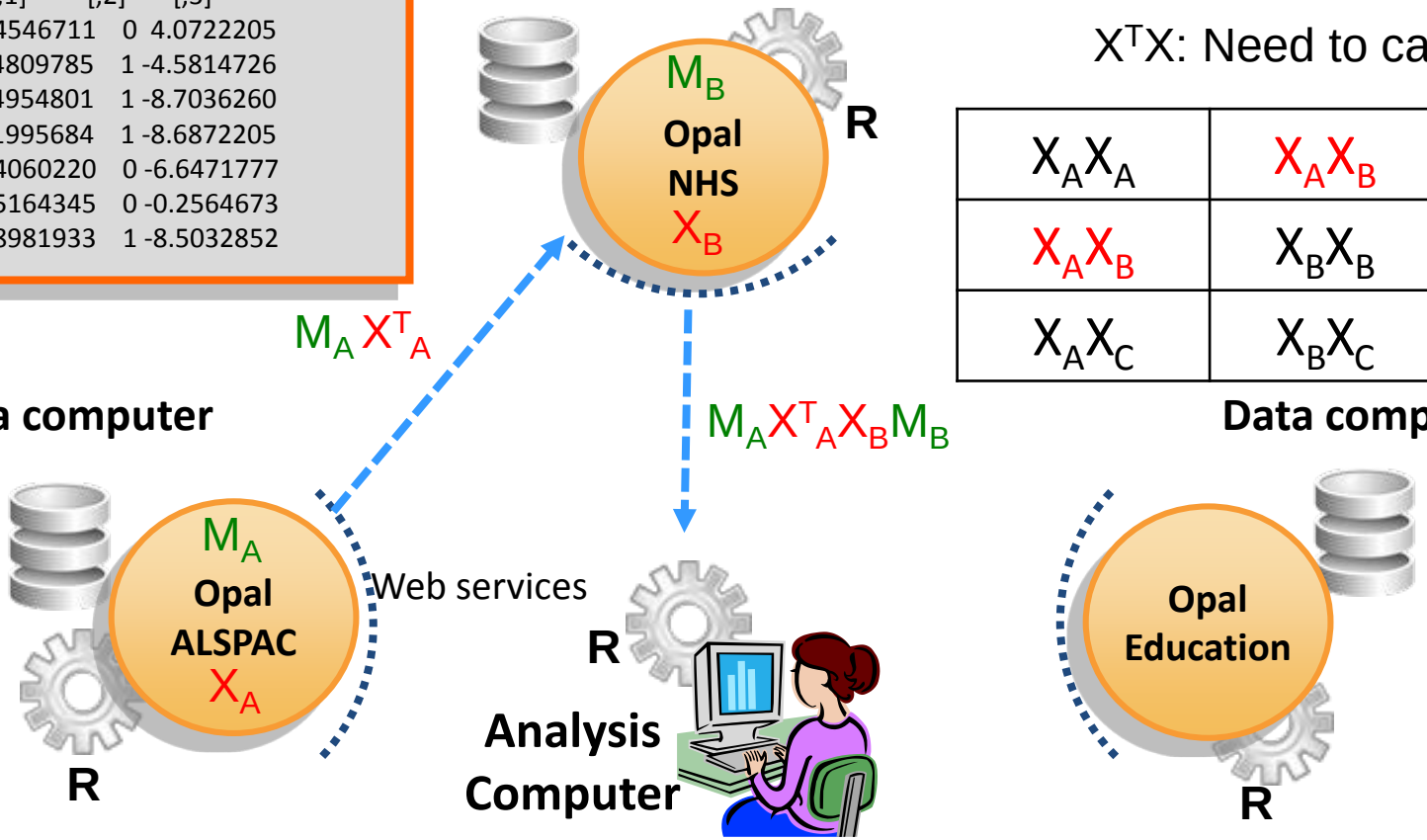
DataSHIELD: current implementation for vertically partitioned (linked) data

```
plain.text.vector.A  plain.text.vector.N
0 1 1 1 0 0 1        1 1 0 1 0 0 1

encryption.matrix
        [,1]       [,2]        [,3]
[1,] -1.444769  2.495677 -5.322736
[2,] -1.355529 -9.369041  2.687347
[3,]  4.603762 -3.622044 -2.817478

occluded.matrix.A
        [,1]   [,2]      [,3]
[1,] -1.4546711   0  4.0722205
[2,]  6.4809785   1 -4.5814726
[3,]  4.4954801   1 -8.7036260
[4,]  0.1995684   1 -8.6872205
[5,] -6.4060220   0 -6.6471777
[6,] -0.5164345   0 -0.2564673
[7,] -5.8981933   1 -8.5032852
```

Regression coefficients = X^TY / X^TX

X^TX: Need to calculate

X_AX_A	X_AX_B	X_AX_C
X_AX_B	X_BX_B	X_BX_C
X_AX_C	X_BX_C	X_CX_C



$$(M_A)^{-1} M_A X^T_A X_B M_B (M_B)^{-1} = X_A X_B$$

SEVENTH FRAMEWORK · wellcome · MRC Medical Research Council · E·S·R·C Economic & Social Research Council · METADAC · DataSHIELD Secure Bioscience Collaboration · D2K data to knowledge

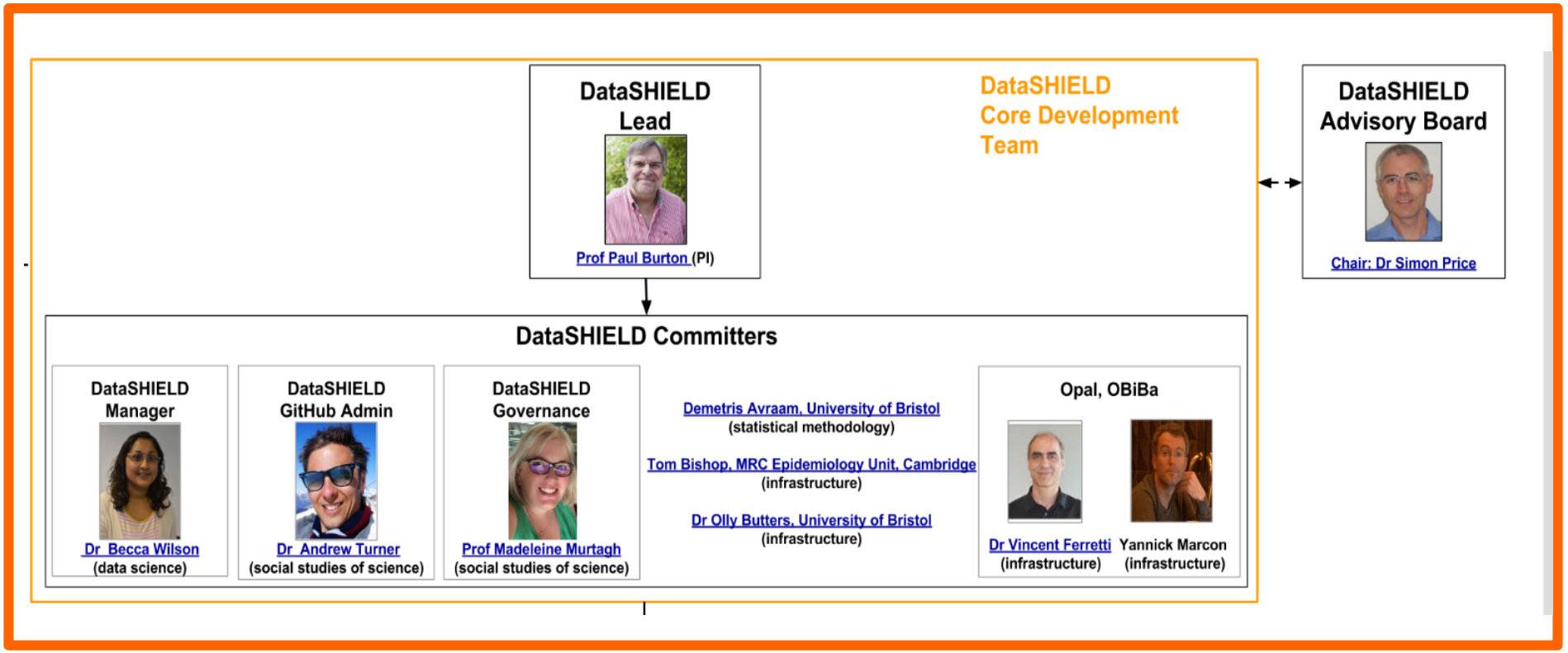
Why DataSHIELD?

- **Horizontal DataSHIELD** multi-site
  - Secured IPD meta-analysis or study-level meta-analysis where different studies hold the same variables on different individuals
  - Data remains behind firewall of study holding data and is invisible and unobtainable externally. Appropriate disclosure settings under control of data controller. 'Local' study data storage could be at a national repository
  - Open-source freeware
- **Vertical DataSHIELD**
  - Ultra-secure analysis of very sensitive linked data where no source is prepared for its linked data to be held by any other sources or a trusted third party
  - Securing the linkage process itself
  - Non-linkage applications eg in 'omics (GCTA, kinship relationship matrix)
- **Single site Horizontal DataSHIELD (a freeware-based data enclave)**
  - Post-publication "open access" to sensitive data
  - "Open access" to simple descriptive stats from rigorously governanced studies
  - Analytic access to sensitive (but not ultra-sensitive) linked data
  - Analytic access to data collected by researchers in resource-poor regions

D2K data to knowledge

SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration

The core DataSHIELD Development Team



SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration

D2K data to knowledge

THANK YOU FOR LISTENING


SEVENTH FRAMEWORK

wellcome

MRC
Medical
Research
Council

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

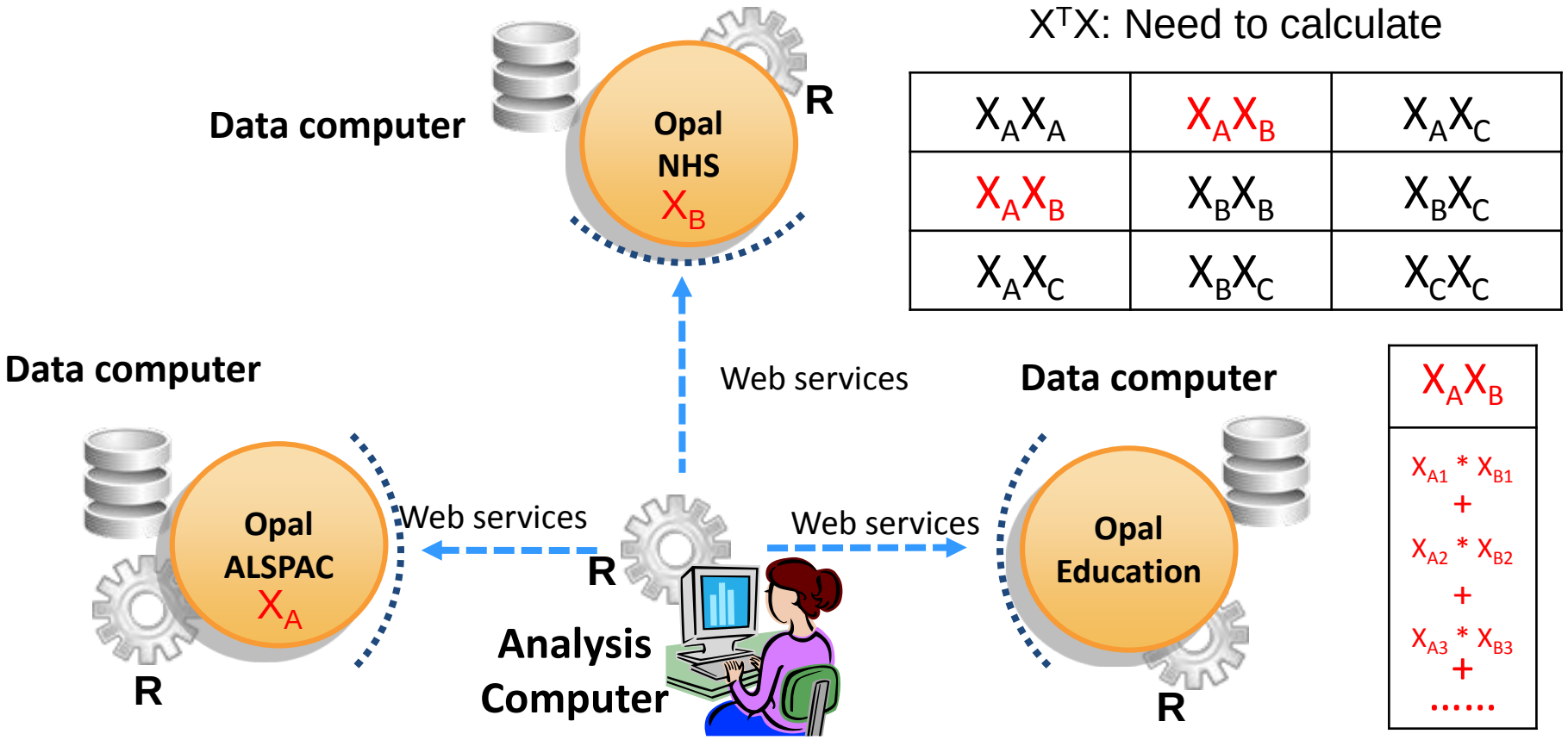
METADAC

DataSHIELD
Secure Bioscience Collaboration

D2K
data to knowledge

DataSHIELD: current implementation for vertically partitioned (linked) data

Regression coefficients = X^TY / X^TX



X^TX: Need to calculate

X_AX_A	X_AX_B	X_AX_C
X_AX_B	X_BX_B	X_BX_C
X_AX_C	X_BX_C	X_CX_C

X_AX_B
$X_{A1} * X_{B1}$ + $X_{A2} * X_{B2}$ + $X_{A3} * X_{B3}$ + ...

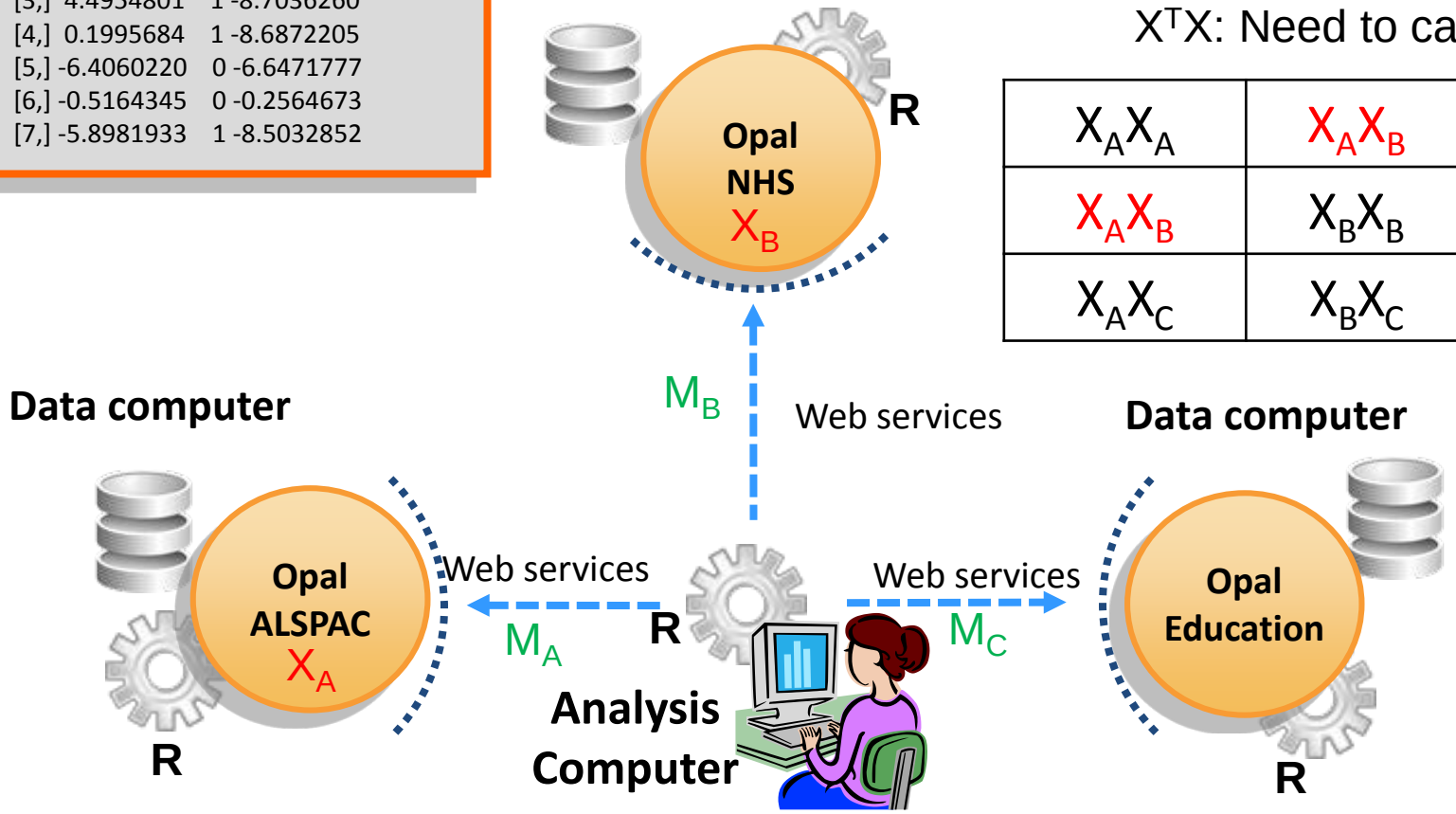
```
plain.text.vector.A  plain.text.vector.N
0 1 1 1 0 0 1        1 1 0 1 0 0 1

encryption.matrix
        [,1]       [,2]          [,3]
[1,] -1.444769  2.495677 -5.322736
[2,] -1.355529 -9.369041  2.687347
[3,]  4.603762 -3.622044 -2.817478

occluded.matrix.A
        [,1]       [,2]      [,3]
[1,] -1.4546711    0  4.0722205
[2,]  6.4809785    1 -4.5814726
[3,]  4.4954801    1 -8.7036260
[4,]  0.1995684    1 -8.6872205
[5,] -6.4060220    0 -6.6471777
[6,] -0.5164345    0 -0.2564673
[7,] -5.8981933    1 -8.5032852
```

DataSHIELD: current implementation for vertically partitioned (linked) data

Regression coefficients = X^TY / X^TX

X^TX: Need to calculate

X_AX_A	X_AX_B	X_AX_C
X_AX_B	X_BX_B	X_BX_C
X_AX_C	X_BX_C	X_CX_C



X_AX_B
$X_{A1} * X_{B1}$ + $X_{A2} * X_{B2}$ + $X_{A3} * X_{B3}$ + ...

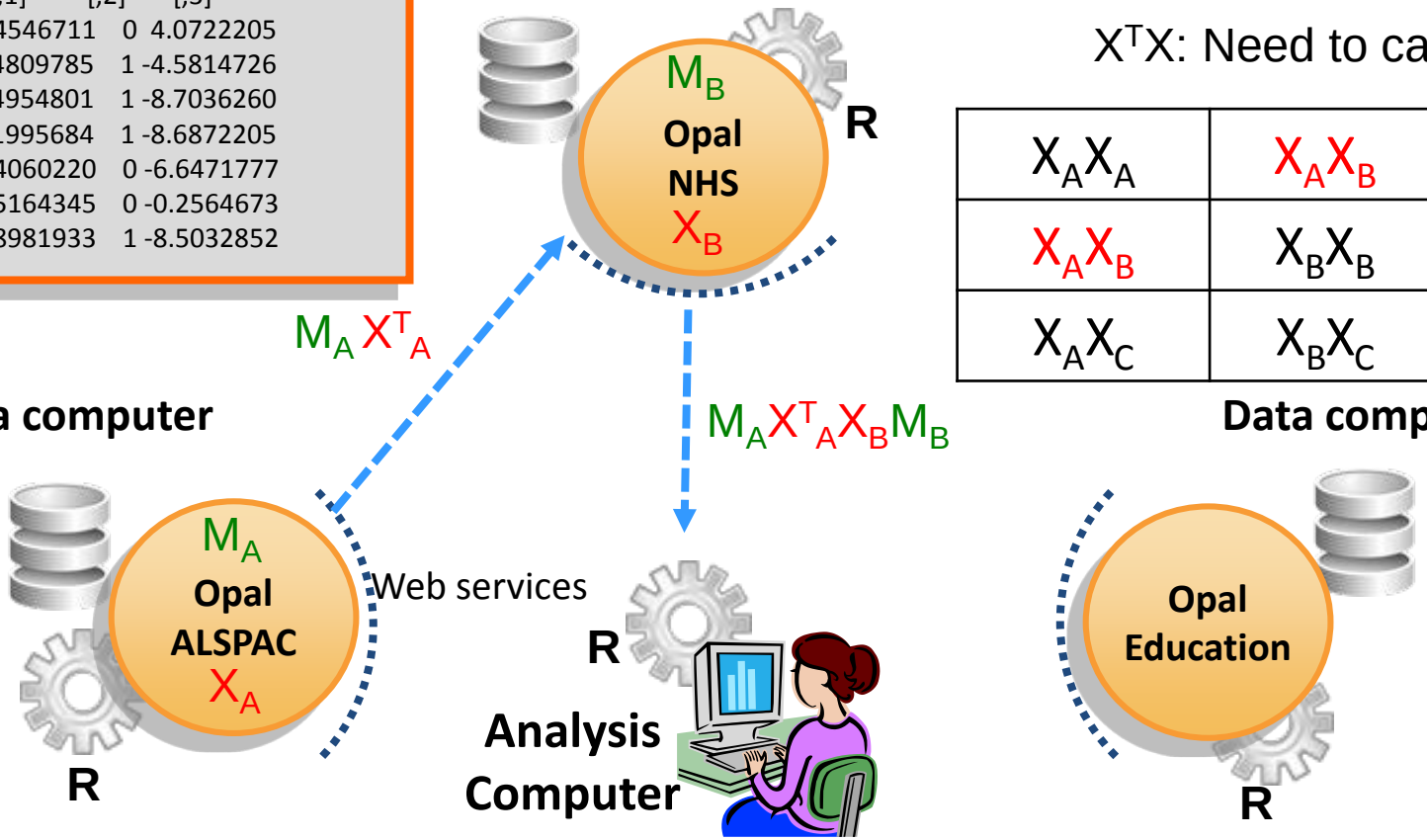
DataSHIELD: current implementation for vertically partitioned (linked) data

```
plain.text.vector.A  plain.text.vector.N
0 1 1 1 0 0 1        1 1 0 1 0 0 1

encryption.matrix
        [,1]       [,2]        [,3]
[1,] -1.444769  2.495677 -5.322736
[2,] -1.355529 -9.369041  2.687347
[3,]  4.603762 -3.622044 -2.817478

occluded.matrix.A
        [,1]   [,2]      [,3]
[1,] -1.4546711   0  4.0722205
[2,]  6.4809785   1 -4.5814726
[3,]  4.4954801   1 -8.7036260
[4,]  0.1995684   1 -8.6872205
[5,] -6.4060220   0 -6.6471777
[6,] -0.5164345   0 -0.2564673
[7,] -5.8981933   1 -8.5032852
```

Regression coefficients = X^TY / X^TX

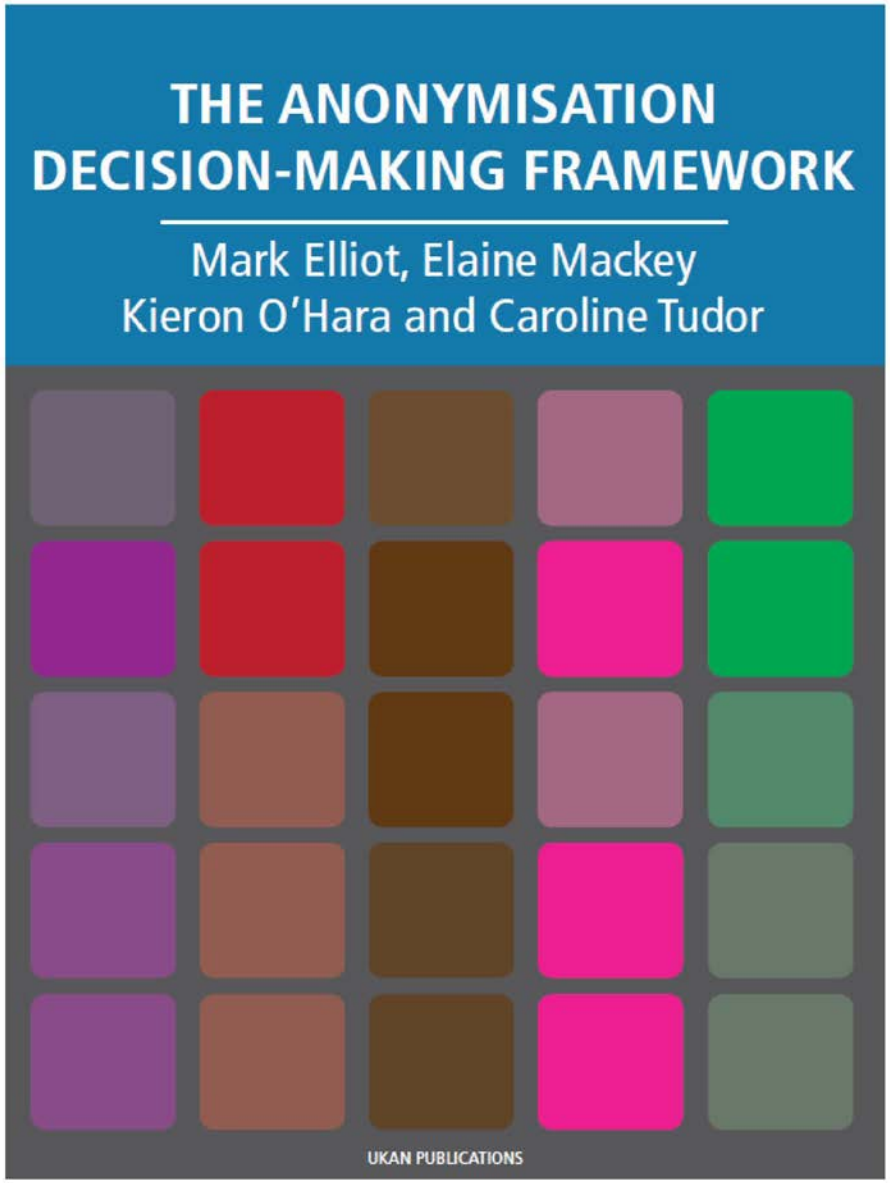
X^TX: Need to calculate

X_AX_A	X_AX_B	X_AX_C
X_AX_B	X_BX_B	X_BX_C
X_AX_C	X_BX_C	X_CX_C



$$(M_A)^{-1} M_A X^T_A X_B M_B (M_B)^{-1} = X_A X_B$$

SEVENTH FRAMEWORK · wellcome · MRC Medical Research Council · E·S·R·C Economic & Social Research Council · METADAC · DataSHIELD Secure Bioscience Collaboration · D2K data to knowledge



Published in the UK in 2016 by
UKAN
University of Manchester
Oxford Road
Manchester
M13 9PL

From Preface:
"Our view has always been that [security] is a heavily context-dependent process and only by considering the data and its environment as a total system (which we call the *data situation*), can one come to a well informed decision about whether and what [approach] is needed."

D2K data to knowledge

SEVENTH FRAMEWORK · wellcome · MRC Medical Research Council · E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL · METADAC · DataSHIELD Secure Bioscience Collaboration

How does matrix-based encryption work?

```
> plain.text.vector.L
[1] 0 1 1 1 0 0 1
> plain.text.vector.N
[1] 1 1 0 1 0 0 1
> sum(plain.text.vector.L*plain.text.matrix.N)
[1] 3
>t(matrix(plain.text.vector.L))%*%matrix(plain.text.vector.N)
     [,1]
[1,]  3
```

How does matrix-based encryption work?

```
> occluded.matrix.L
          [,1]        [,2]      [,3]
[1,] -1.4546711    0  4.0722205
[2,]  6.4809785    1 -4.5814726
[3,]  4.4954801    1 -8.7036260
[4,]  0.1995684    1 -8.6872205
[5,] -6.4060220    0 -6.6471777
[6,] -0.5164345    0 -0.2564673
[7,] -5.8981933    1 -8.5032852
> e.mat.L
          [,1]        [,2]         [,3]
[1,] -1.444769  2.495677 -5.322736
[2,] -1.355529 -9.369041  2.687347
[3,]  4.603762 -3.622044 -2.817478
> e.mat.L%*%occluded.matrix.L
          [,1]        [,2]         [,3]         [,4]         [,5]         [,6]         [,7]
[1,] -19.57369  17.51813  42.32785  48.44713  44.636397  2.11123627  56.277949
[2,]  12.91532 -30.46620 -38.85246 -32.98514  -9.179719  0.01082581 -24.225142
[3,] -18.17035  39.12303  41.59635  21.77277 -10.763524 -1.65495077  -6.818104
```

```
> plain.text.vector.L
 [1] 0 1 1 1 0 0 1
```

How does matrix-based encryption work?

```
> e.mat.L%*%occluded.matrix.L
          [,1]       [,2]       [,3]       [,4]       [,5]       [,6]       [,7]
[1,] -19.57369  17.51813  42.32785  48.44713  44.636397  2.11123627  56.277949
[2,]  12.91532 -30.46620 -38.85246 -32.98514  -9.179719  0.01082581 -24.225142
[3,] -18.17035  39.12303  41.59635  21.77277 -10.763524 -1.65495077  -6.818104
> plain.text.vector.N
[1] 1 1 0 1 0 0 1
> e.mat.L%*%occluded.matrix.L%*% plain.text.matrix.N
          [,1]
[1,] 102.66952
[2,] -74.76116
[3,]  35.90735
> inv.e.mat.L%*%e.mat.L%*%occluded.matrix.L%*% plain.text.matrix.N
[,1]
[1,]  -0.6723174
[2,]   3.0000000
[3,] -17.6997578
> sum(plain.text.vector.L*plain.text.matrix.N)
[1] 3
```

SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration

D2K data to knowledge

Why do we need to occlude the original plain text vector?

```
> plain.text.vector.L
[1] 0 1 1 1 0 0 1

> e.mat.1
        [,1]
[1,] 7.13763

> e.mat.1%*%t(matrix(plain.text.vector.L))
     [,1]  [,2]        [,3]        [,4]        [,5] [,6][,7]
[1,]    0 7.13763 7.13763 7.13763    0    0 7.13763

>e.mat.1%*%t(matrix(plain.text.vector.L))%*%plain.text.matrix.N
         [,1]
[1,] 21.41289

>(1/e.mat.1)*e.mat.1%*%t(matrix(plain.text.vector.L))%*%plain.tex
t.matrix.N
     [,1]
[1,]  3
```

D2K data to knowledge

SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration

Promoting effective 'data security, privacy and trust' in our data systems

- A complex challenge involving science, technology, governance and other fundamental social issues
- **True** transdisciplinary programs of work are essential
- No single solution can ever be a 'silver bullet'
- Even the most sophisticated combination of solutions will **never** promise optimally efficient exploitation of available data with **zero** risk of undesirable disclosure events, mistakes in data management and/or of malign attacks on data
- **DataSHIELD can provide one component of an integrated solution to a range of important challenges**

SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration

D2K data to knowledge

Conventional analysis

STUDY	N	LOG-ODDS RATIO	STANDARD ERROR
1958BC	7210	1.7063198	0.16427649
SHIP	4308	0.6368959	0.11976096
PREVEND	8592	0.6888834	0.18821449
LIFELINES	90920	1.4431487	0.04800230
MITCHELSTOWN	2048	0.9541291	0.20801307
FINRISK	5024	1.2253285	0.14526764
CHRIS	1583	1.4502807	0.35001932
MICROS	1060	0.9276257	0.40205915
KORA	3080	1.1576405	0.15144140

Does it work?

Association of diabetes with BMI>30 in 9 HOP studies

DataSHIELD analyses

Individual level meta-analysis

Term	LOR	SE	OR	Low95%CI	Up95%CI
MODEL= "diab ~ S.1 + S.2 + S.3 + S.4 + S.5 + S.6 + S.7 + S.8 + bmi + age + female", logistic regression					
BMI>30	1.292	0.038	3.640	3.379	3.921
AGE	0.0715	0.0016	1.074	1.071	1.078
FEMALE	-0.329	0.0374	0.720	0.669	0.774

Random effects study level meta-analysis

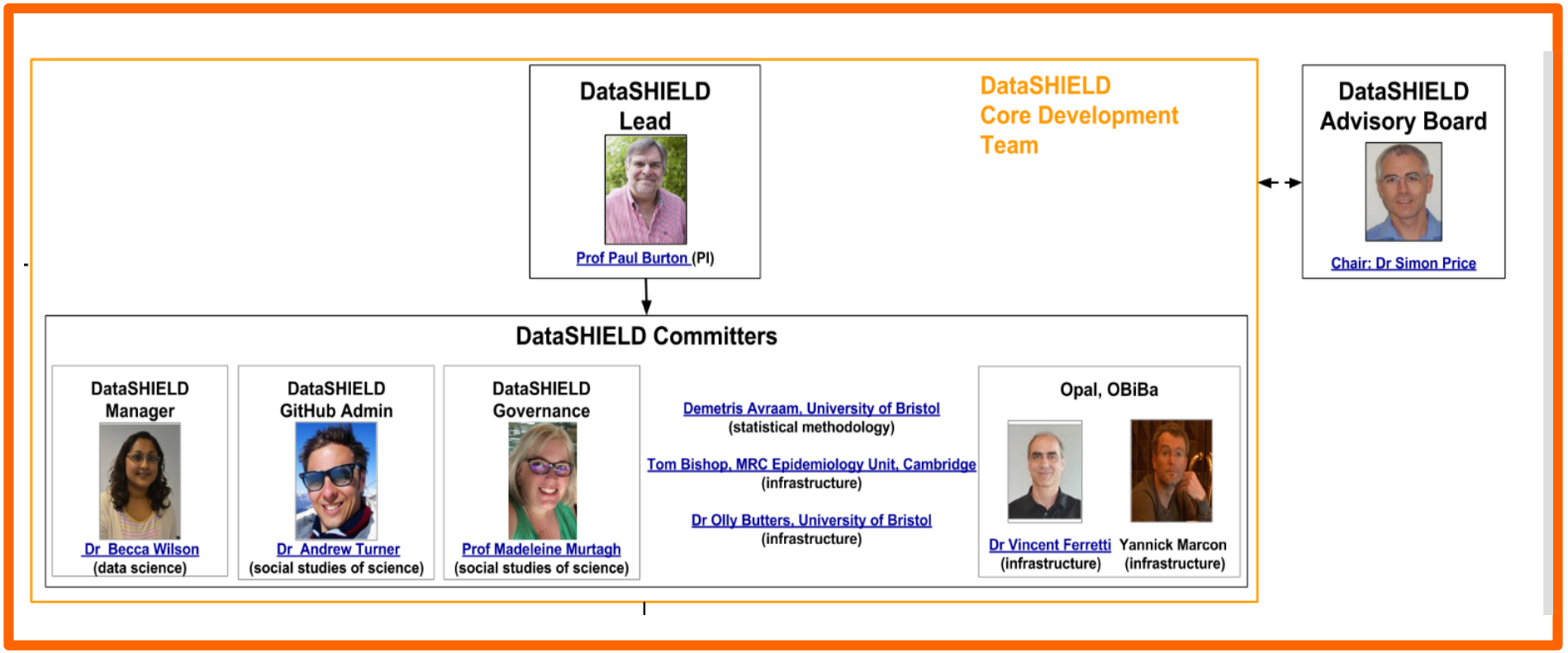
Term	LOR	SE	OR	Low95%CI	Up95%CI
MODEL= "diab ~ bmi + age + female", logistic regression					
BMI>30	1.137	0.129	3.117	2.419	4.016
Test for Heterogeneity: Q(df = 8) = 61.1275, p-val < .0001					

SEVENTH FRAMEWORK

wellcome

D2K data to knowledge

The core DataSHIELD Development Team



SEVENTH FRAMEWORK wellcome MRC Medical Research Council E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL METADAC DataSHIELD Secure Bioscience Collaboration

D2K data to knowledge