Digital Curation for the Public's Health: Ethics, Security & Trust

Digital Curation for the Public's Health

Ethics, Security & Trust

DCC/RIN Research Data Management Forum
"Dealing with Sensitive Data: managing ethics, security and trust"

10th March 2010

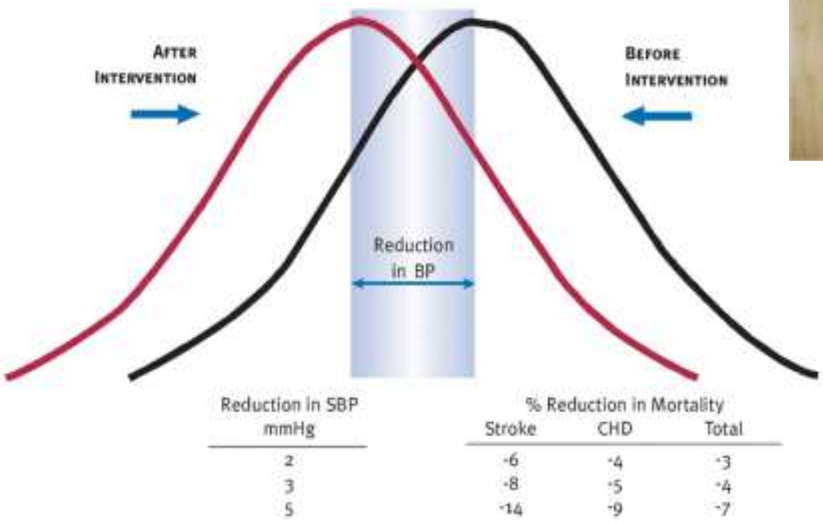
Prof. Iain Buchan
University of Manchester

Digital Curation for the Public's Health

- Where does the public's health need digital innovation?

- How can research curators promote this innovation?
  - Ethics
  - Security
  - Trust

- Is a framework required?
  - Social contract
  - Digital & operational infrastructure: e-Lab



Source: Whelton PK, et al. Primary prevention of hypertension: Clinical and public health advisory from The National High Blood Pressure Education Program. JAMA 2002;288:1882–8.

Public Health Needs for Digital Innovation: I

CITIZEN-LED PREVENTION & EARLY INTERVENTION

Approaches to Healthcare

- <u>Clinical</u> model:
  - Rescue the ill
  - Resource $\propto$ illness
  - Specialise to optimise

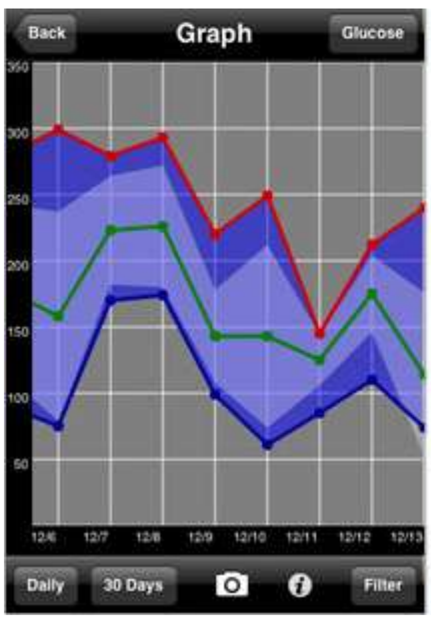
Reactive

- <u>Public Health</u> model:
  - Rescue the at-risk
  - Resource $\propto$ disease/risk
  - Generalise to optimise

Proactive

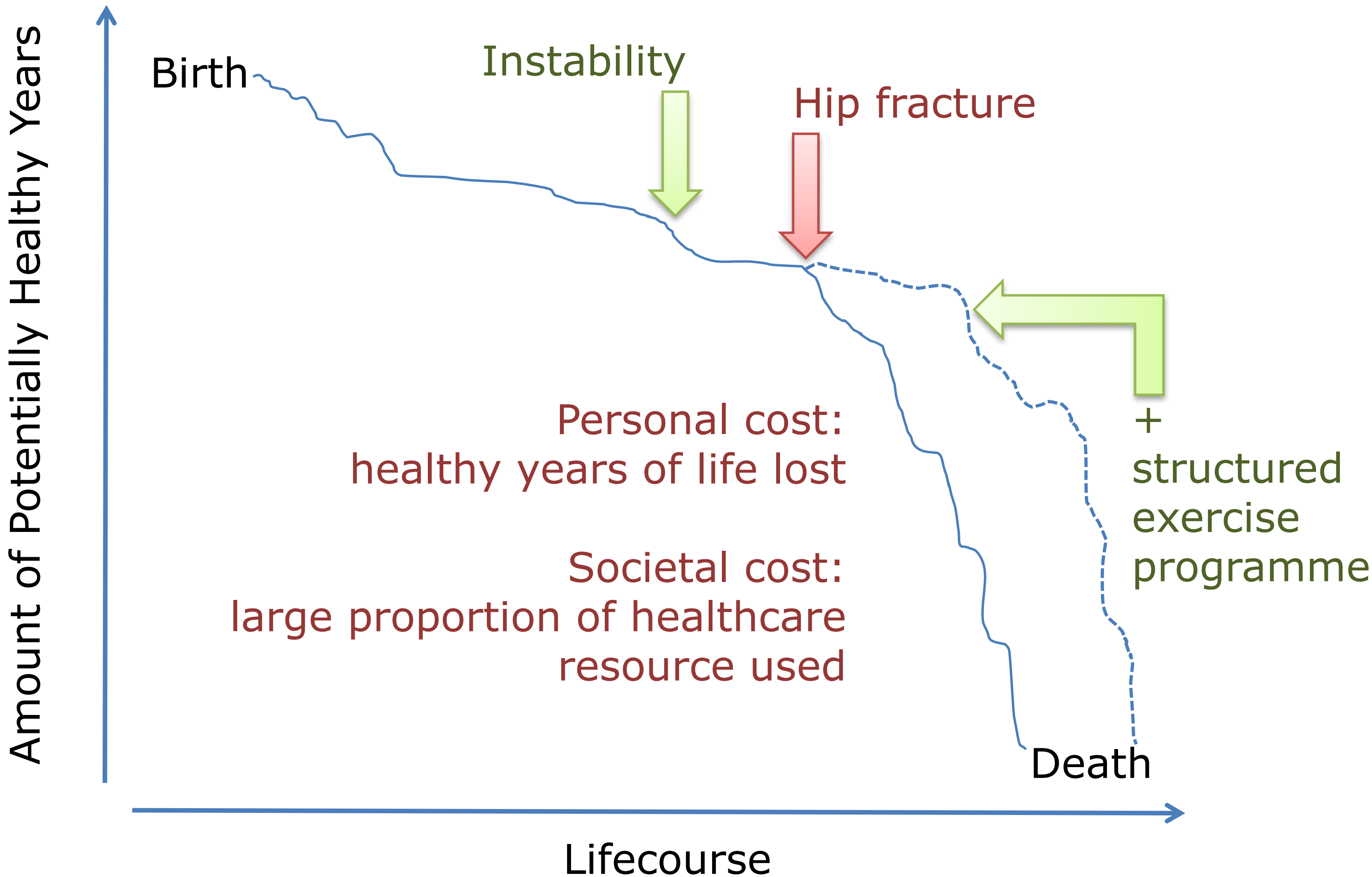
Healthcare Digital Innovation is Mostly Clinical

- Specialist-driven
- Excellent in niches
- Incremental



…while public health technologies may be left to perverse market-forces.

A life-course view of personal health



State-led:
Health Promotion

Citizen-led:
Health & Wellbeing

Manchester Public Health Development Service

GATE 65+

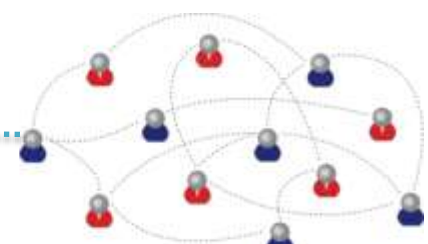
Getting Active Through Exercise

Manchester NHS Primary Care Trust

Know → Act

Persuasive Technology

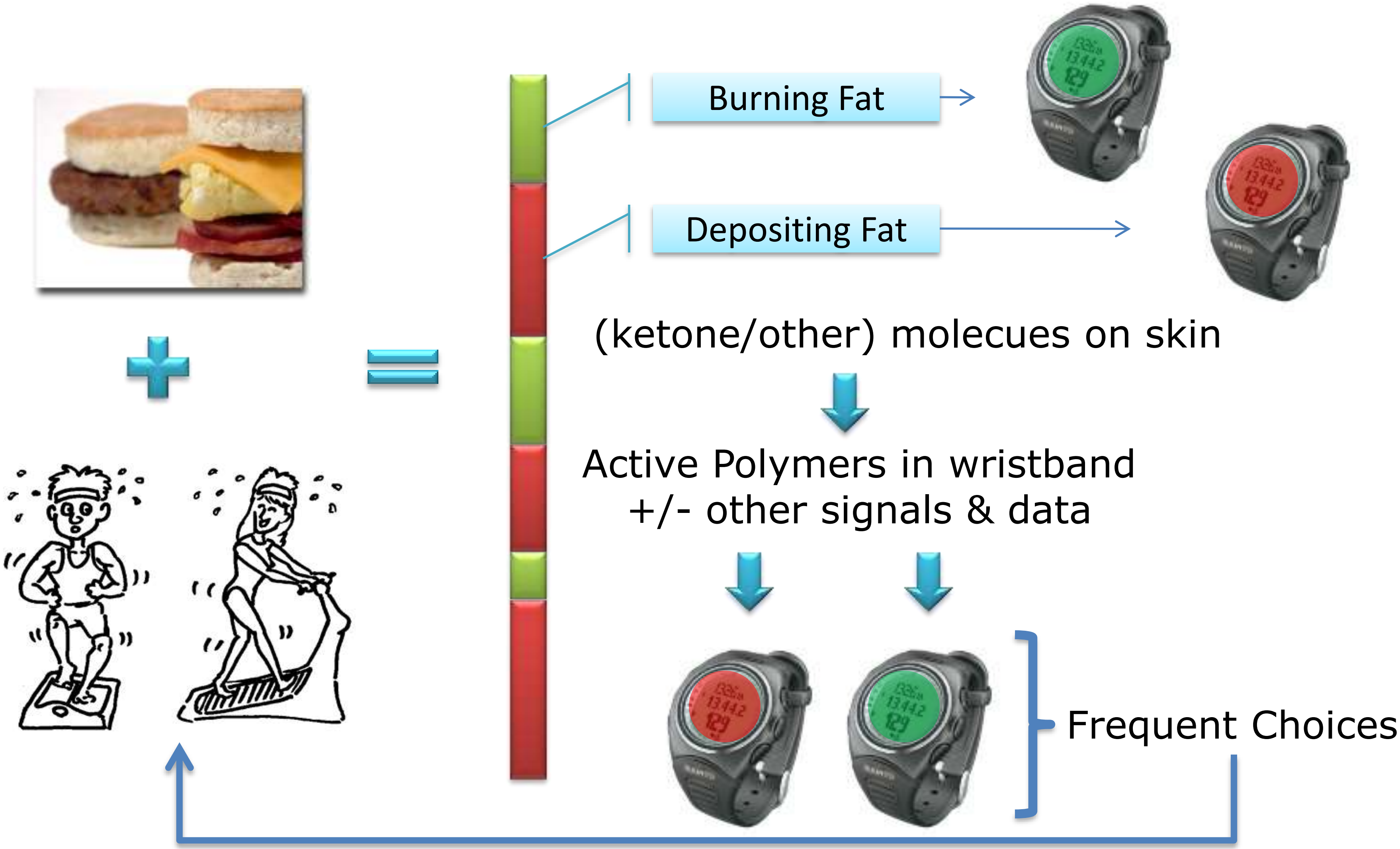
Feel ← Interact



Social Marketing

Social Networks

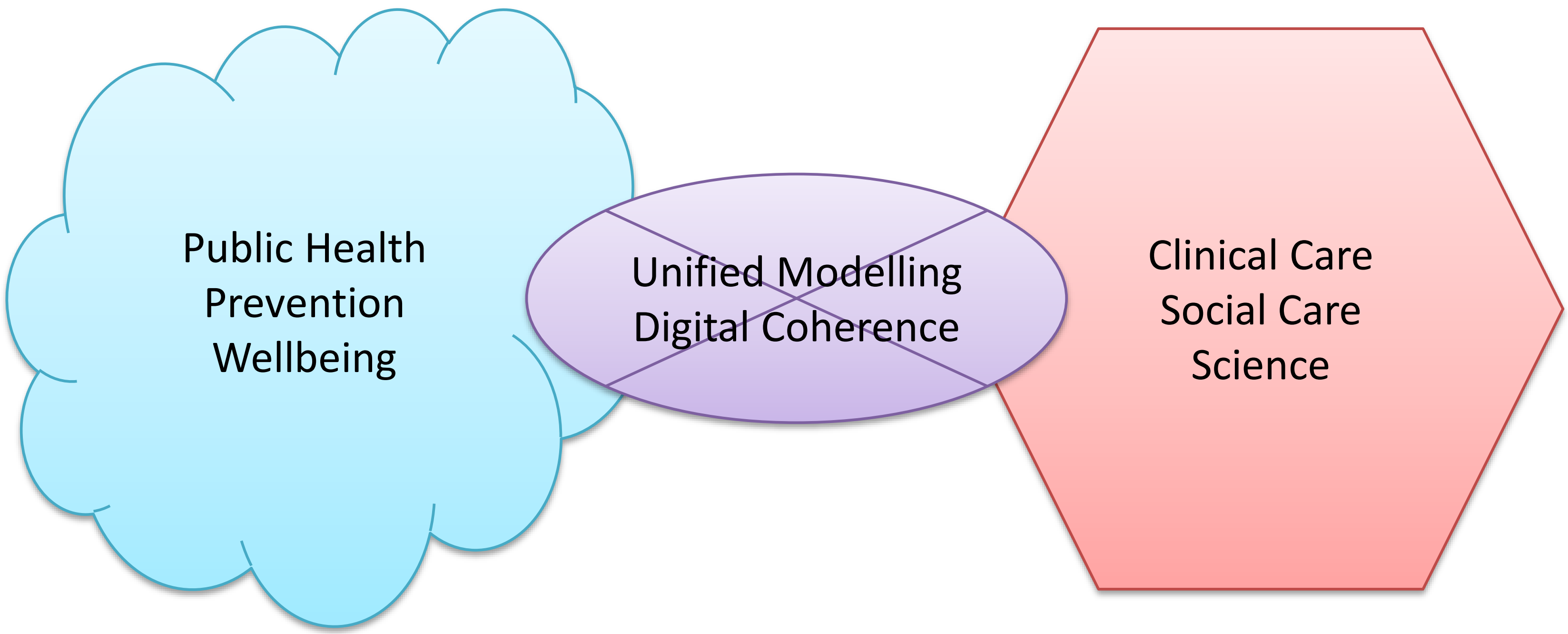
Re-wire the brain to resist over-malnutrition?



Care costs escalate without prevention

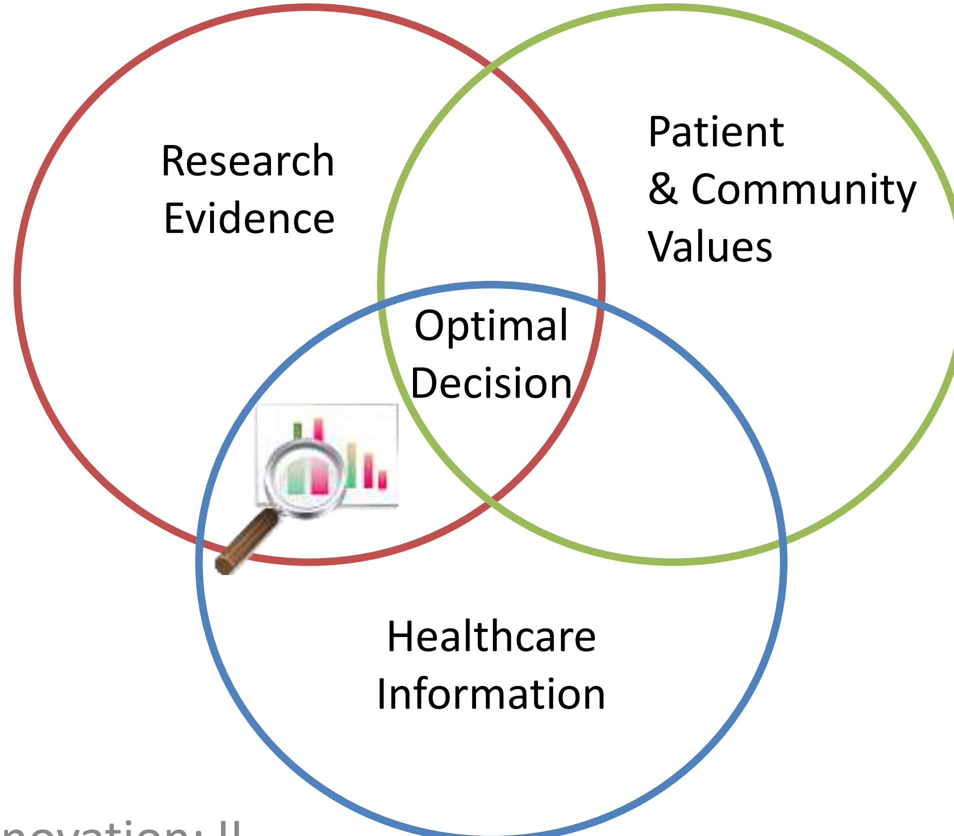
Citizen > State

State > Citizen



Life-course view
(barely visible investment)
Life of a person

Sickness view
(fire-fighting investments)
Life of a parliament



Public Health Needs for Digital Innovation: II

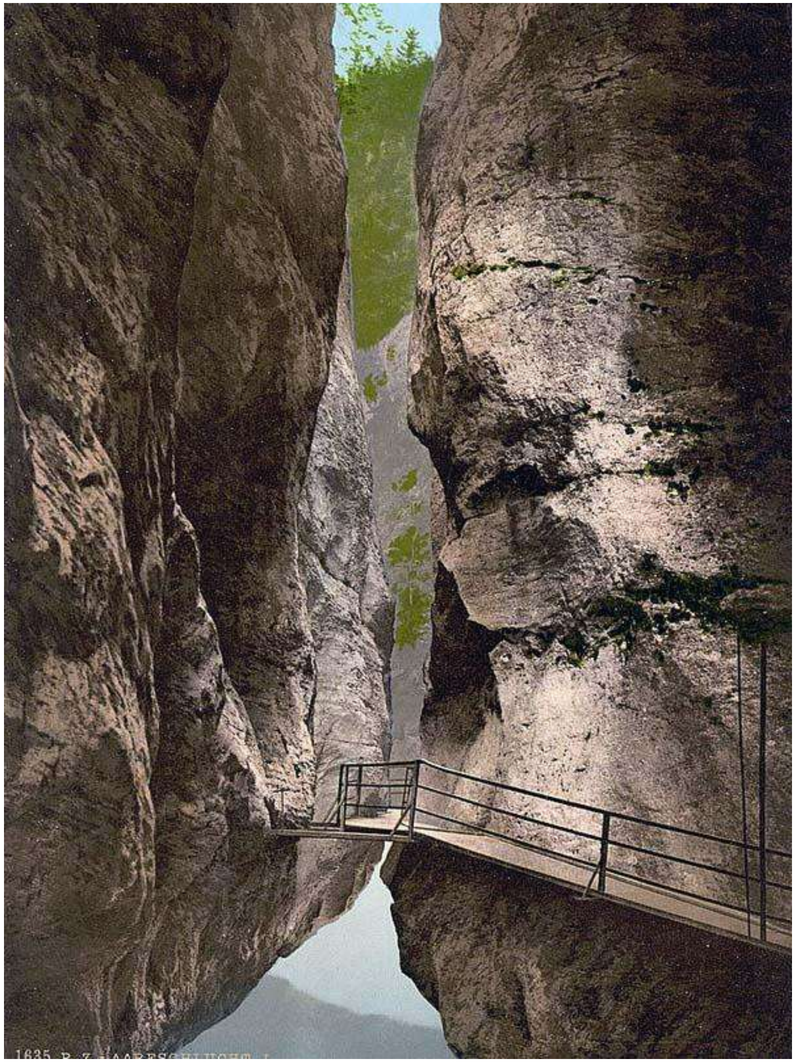
DYNAMIC LINKAGE OF RESEARCH INTO PRACTICE AND PRACTICE INTO RESEARCH

Healthcare Problem: **Gaps** in Communication & Organisation

Self Care

Primary Care

Hospital A



Clinical Care

Secondary Care

Hospital B

Digital Bridges Since 1990s:
Integrated Care **Pathways (Disease-specific)**

Self Care

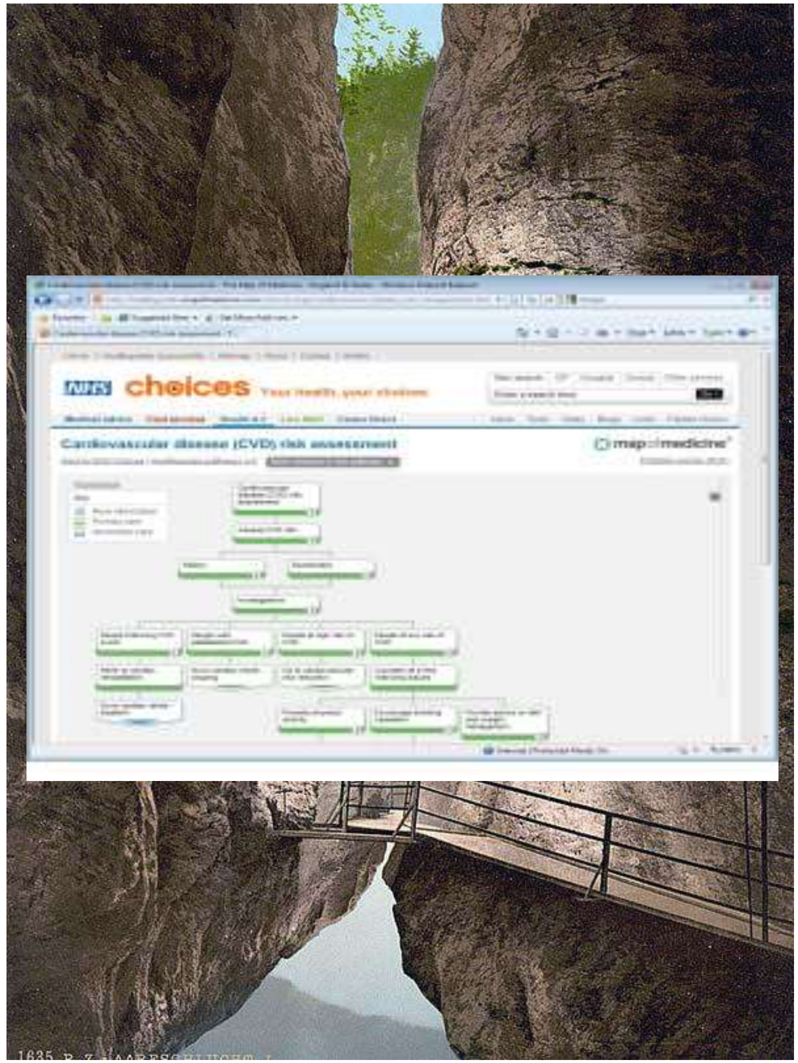
Clinical Care

Primary Care

Secondary Care

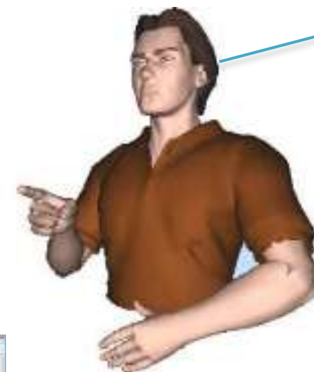
Hospital A

Hospital B



Missing: Patient & Community
‘**Big-picture**’ Across Disease / Specialist Pathways

Person:
Experiencing n of 1 trials

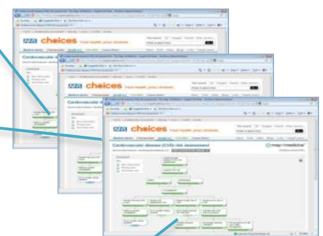


Self Care

Clinical Care

Diabetology:
Glucose control

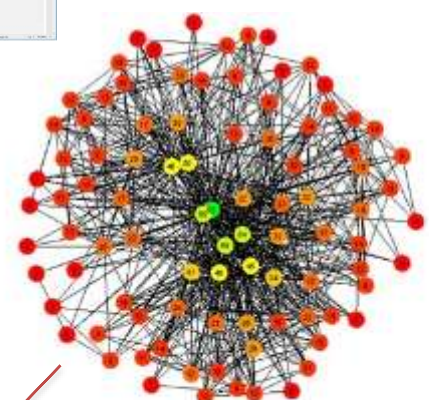
Ophthalmology:
Diabetic eye care



∪ = Avatar

Secondary Care

Primary Care

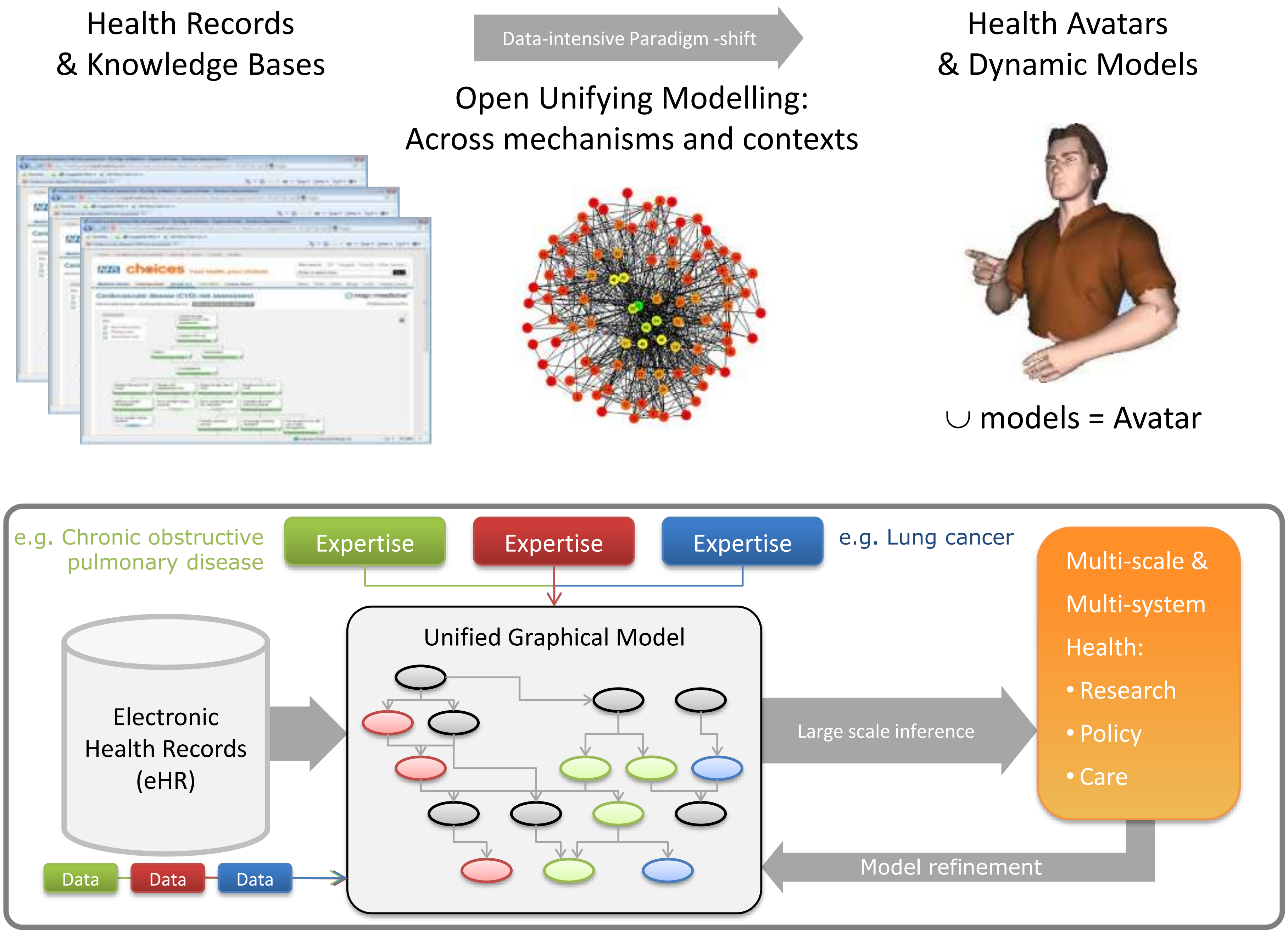


Nephrology:
Chronic kidney disease

Hospital A

Hospital B

Future Health ‘Avatar’:
Fuse research evidence about the average person
with a longitudinal model of the person represented



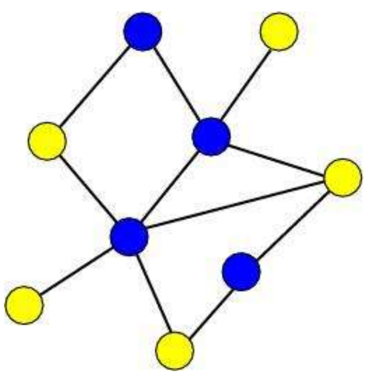
Health Records
& Knowledge Bases
Data-intensive Paradigm -shift
Health Avatars
& Dynamic Models
Open Unifying Modelling:
Across mechanisms and contexts
∪ models = Avatar
e.g. Chronic obstructive pulmonary disease
Expertise
Expertise
Expertise
e.g. Lung cancer
Multi-scale &
Multi-system
Health:
• Research
• Policy
• Care
Electronic Health Records (eHR)
Unified Graphical Model
Large scale inference
Model refinement
Data
Data
Data

Trans-disciplinary Analysis of Clinical Research Data

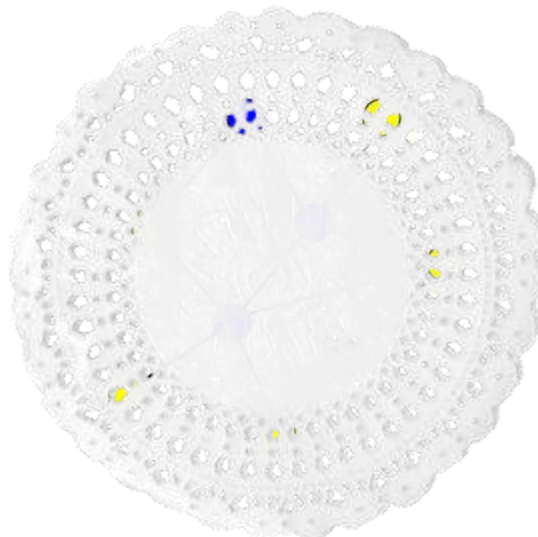
Machine Learning in Epidemiology...

Introducing Health Sciences Signal Paths to Physical Scientists & Engineers

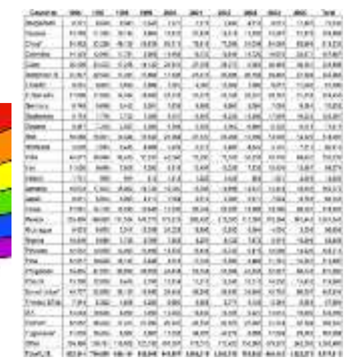
Problem Space



Observation Space



Data Space



$y = b1x1 + b2x2 + b3x3 + c$

...like resolving an image through a prism through a doyley

Hypothesis-driven Epidemiology: Sieving Associations

Association	**Bias**	**Type**	**Explanation**
C→M	Cause-effect	Real	Cause-effect
MI→C	Reverse	Real	Effect-cause
C←?→MI	Confounding	Real	Effect-effect
C </> MI	Random error	Spurious	Chance
C </> MI	Systematic error	Spurious	Bias

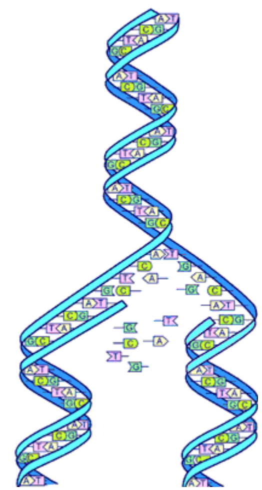
C = caffeine, MI = myocardial infarction (heart attack)

Disciplined approach to causal inference, Bradford-Hill: Criteria (temporality, strength, dose-response, consistency, plausibility, consideration of alternatives, open to experiment, specificity, coherence)

Need to Address Complexity & Scale

Problem 1:
Dwindling hits from tools to detect independent "causes"

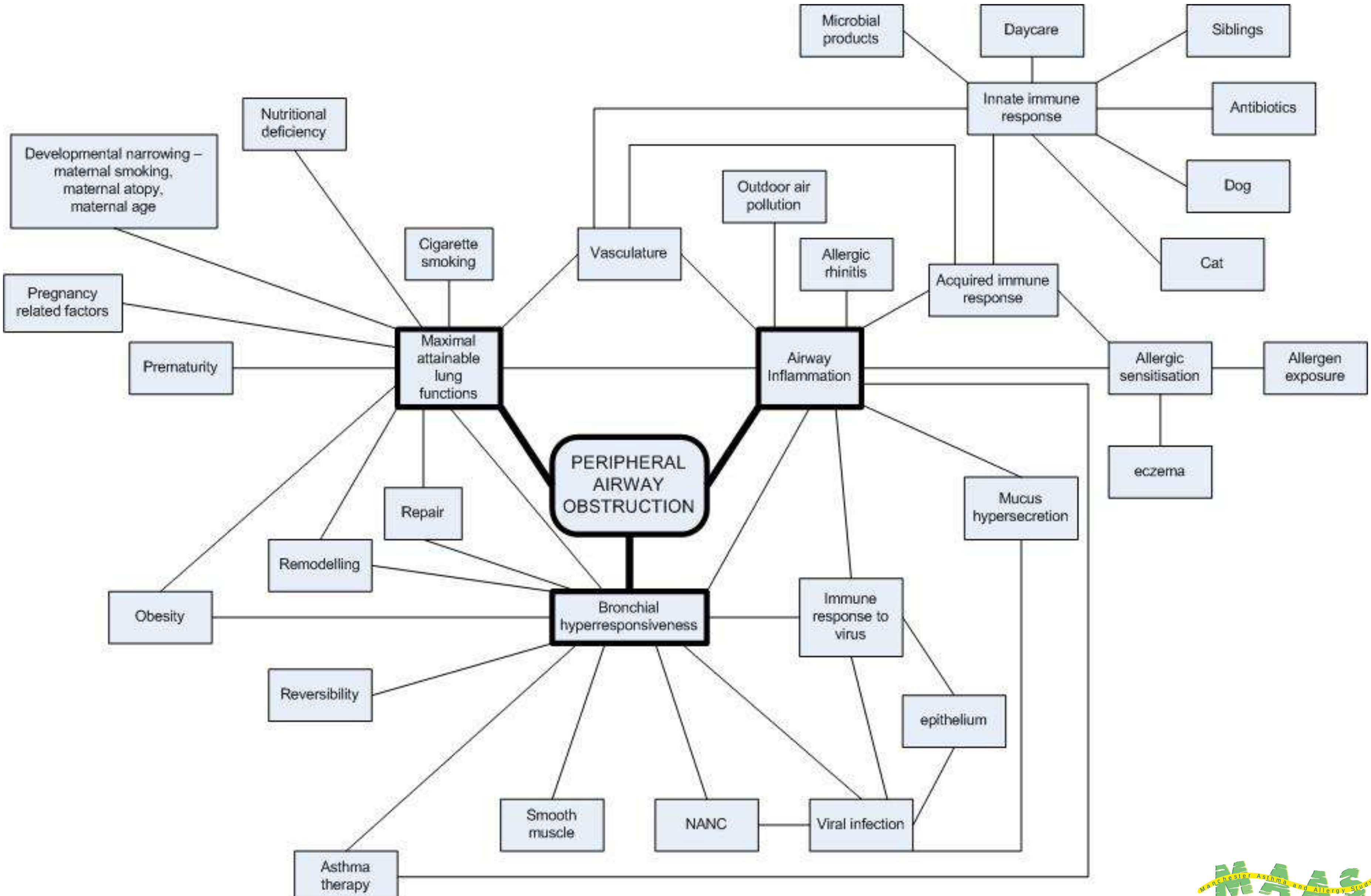
Problem 2:
Knowledge can't be managed by reading papers any more



The big public health problems
e.g. Type 2 Diabetes
have "complex webs of causes"

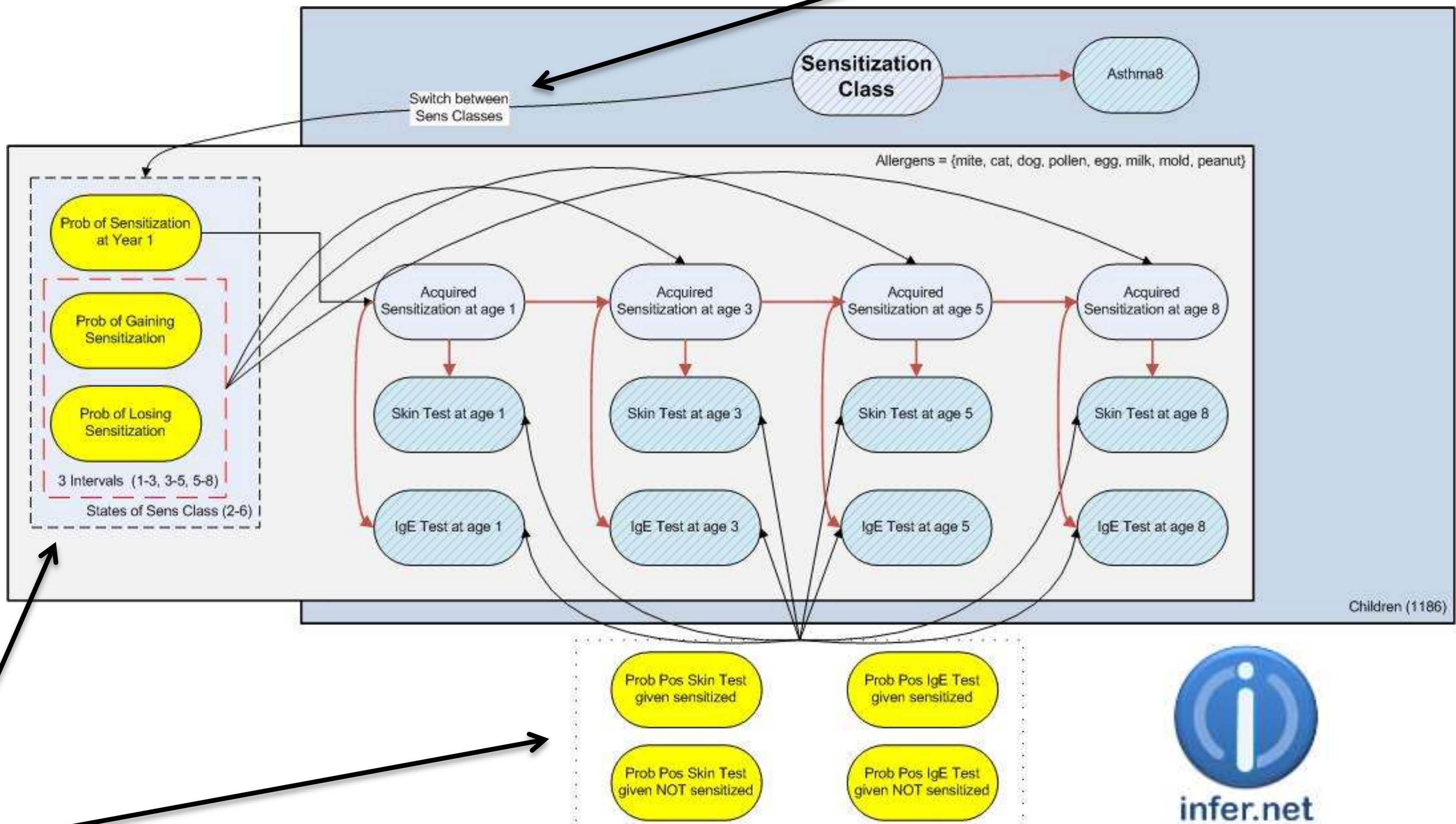
The "data-set" and structure
extend beyond
the study's observations

A Graphical Model of Asthma



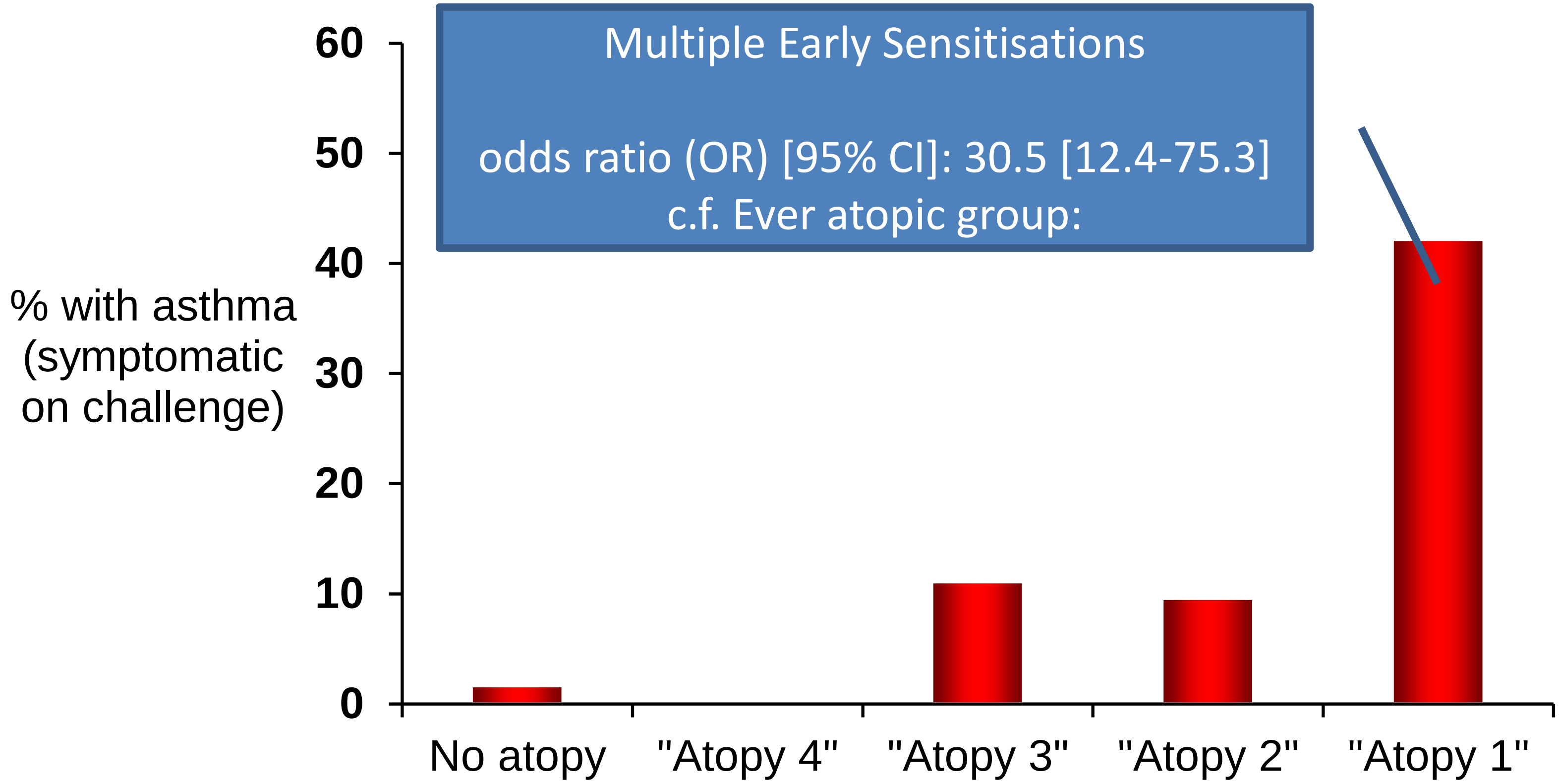
The Atopy Model

Unsupervised Clustering



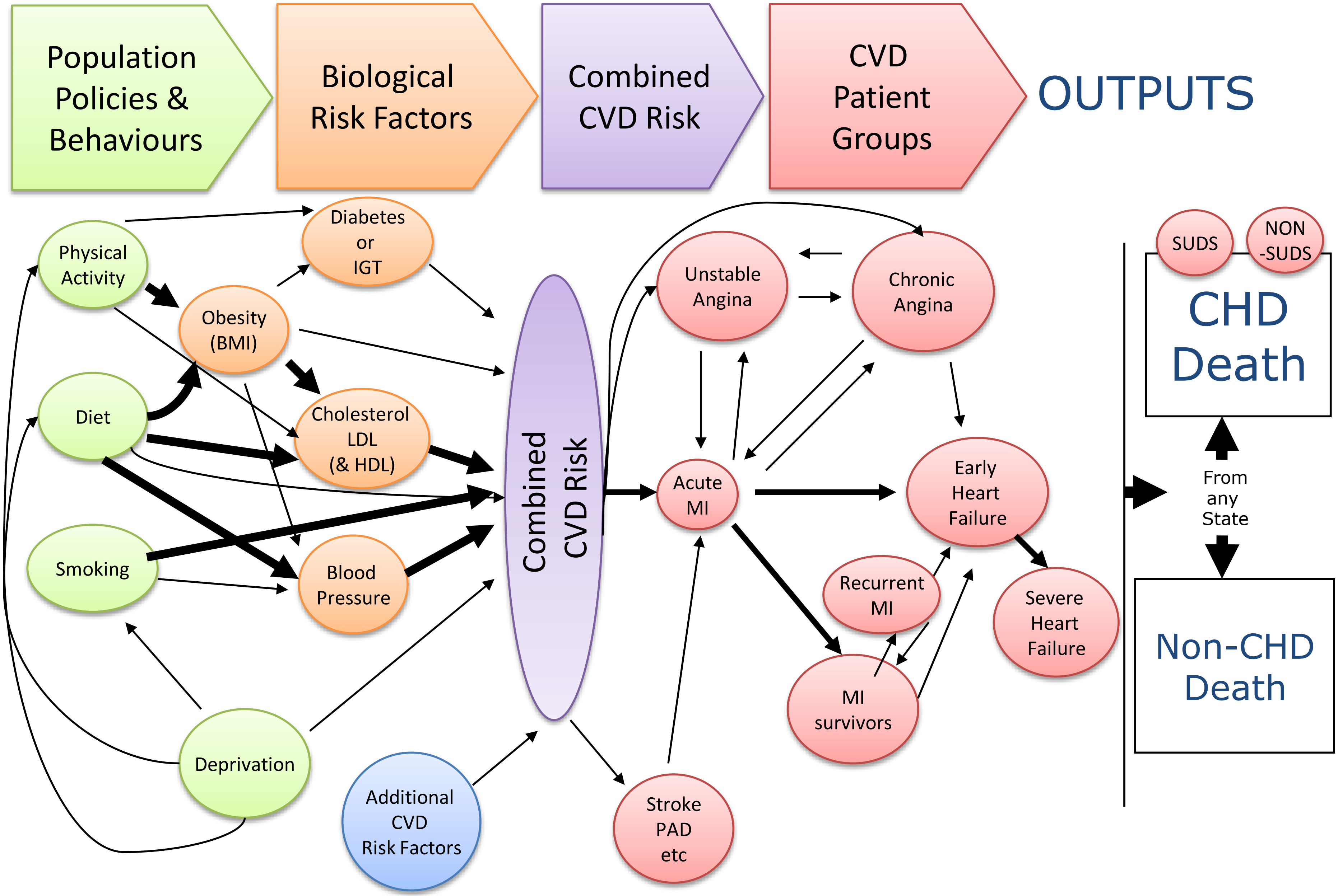
To Infer

Atopy Classes and Asthma



Trans-disciplinary Simulation of Public Health Impacts

Digital Assets → Policy Decisions...

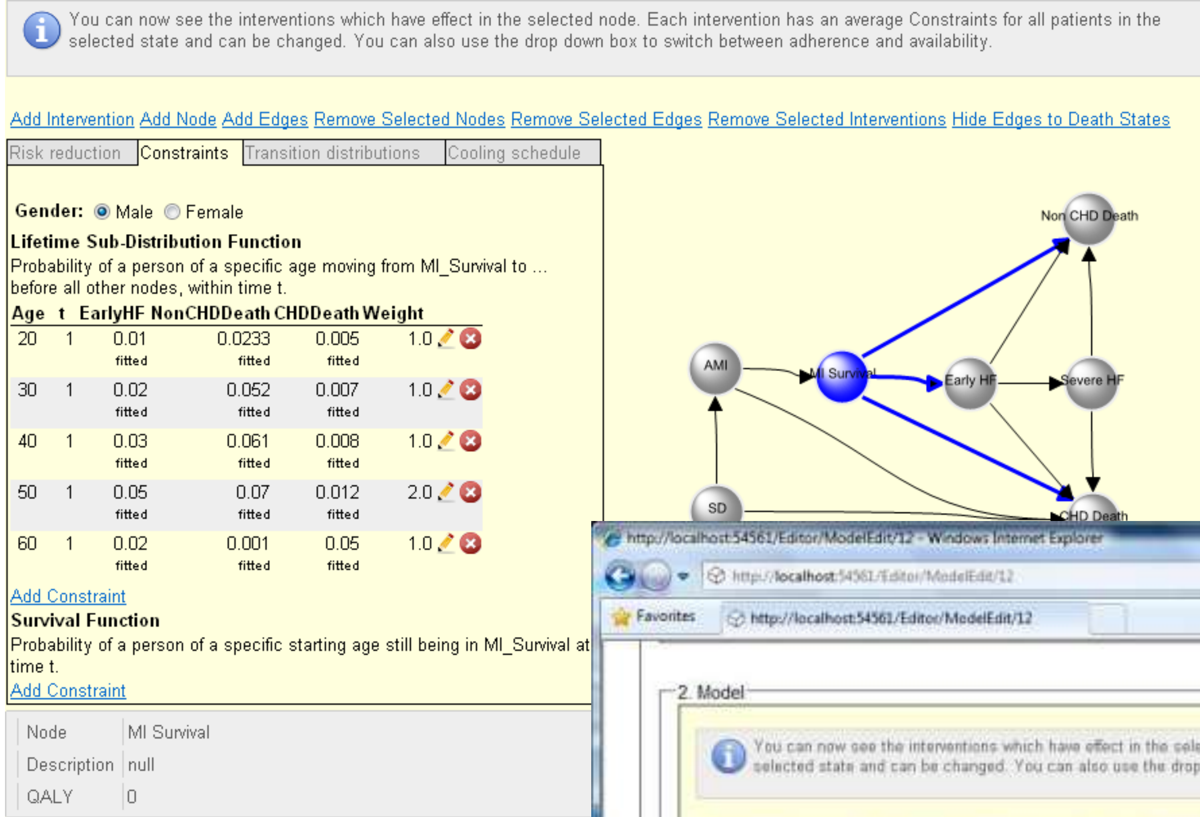


Pulling evidence together into one, realistically-complex model: e.g. MRC IMPACT II

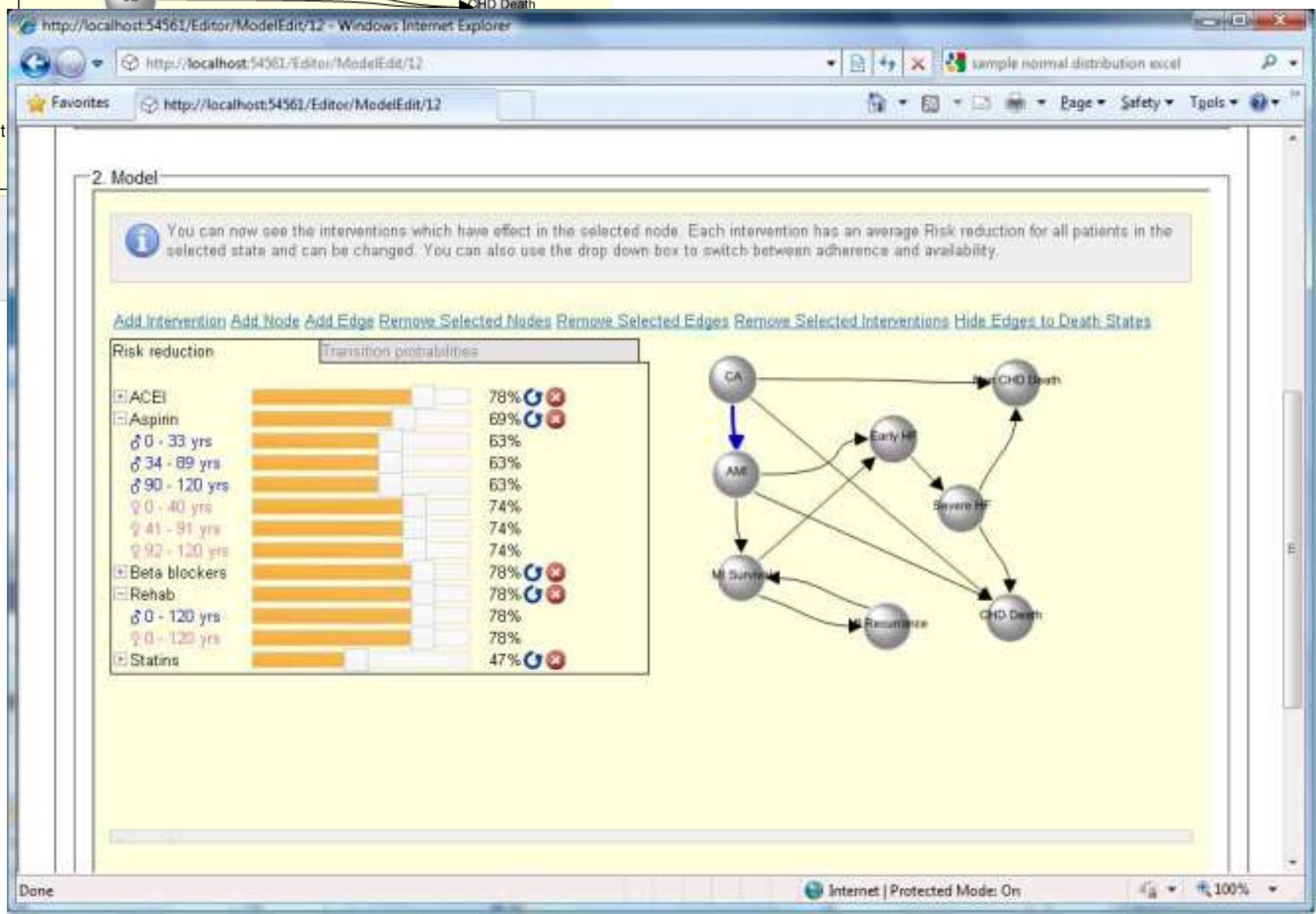
Outputs: Population-based incidence, prevalence; Deaths prevented; Life-Years; Life expectancy; Costs; Cost-effectiveness ratios

Medical Research Council MRC

NHS National Institute for Health Research

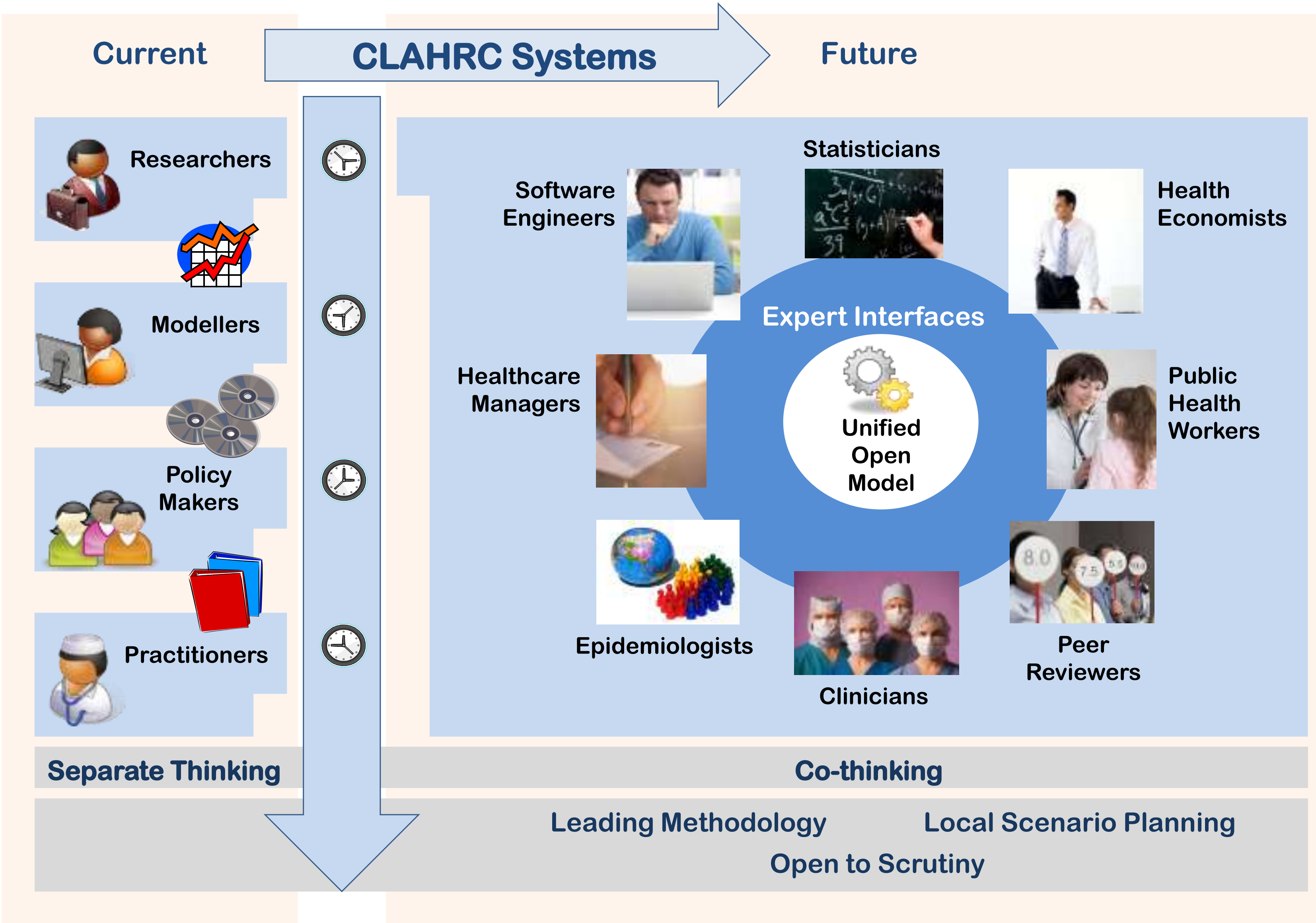


[Recombine evidence around probabilistic graphical models of disease & care-services]

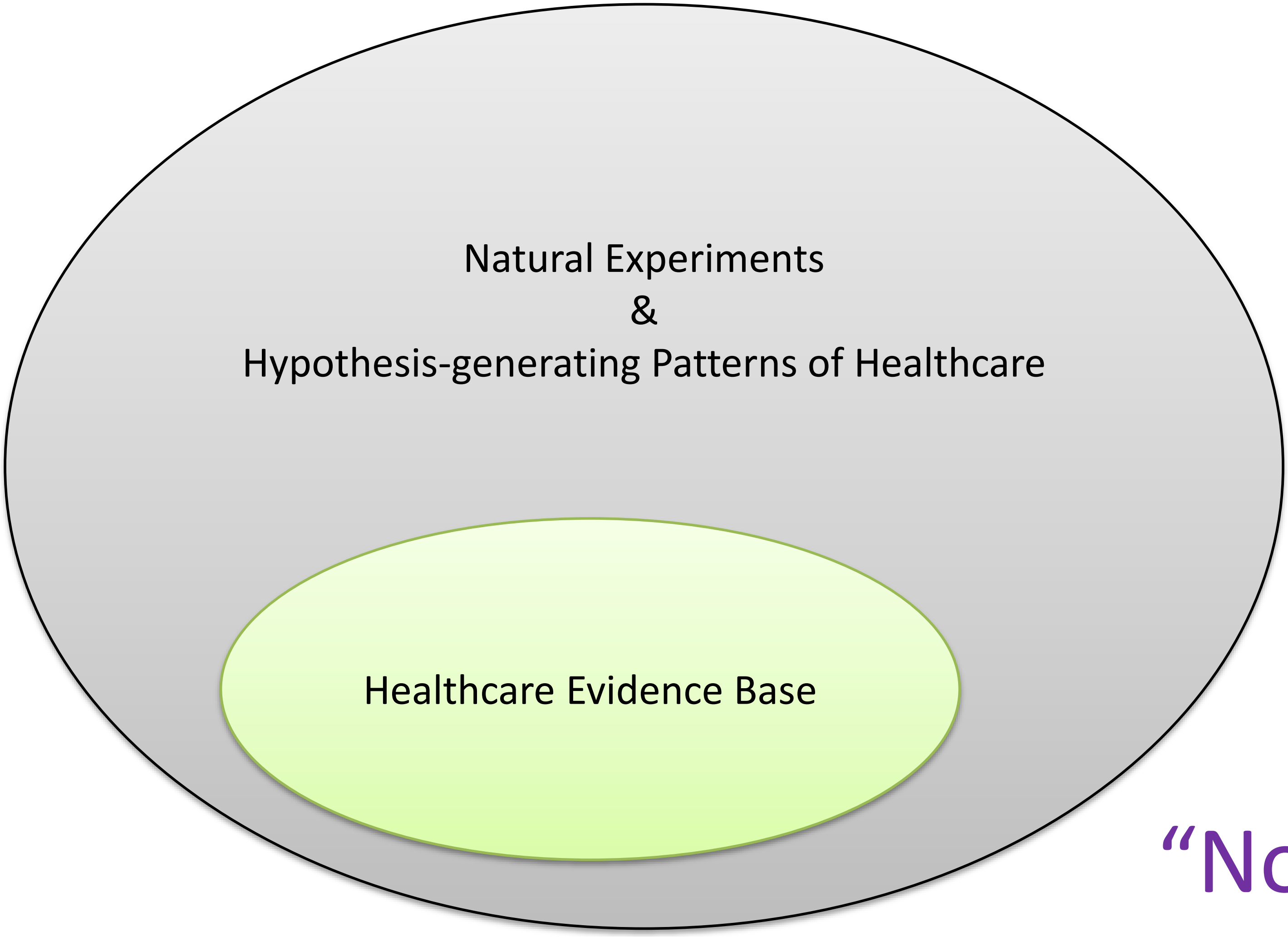


Ask 'what if'
invest in statins
vs. smoking cessation
etc.

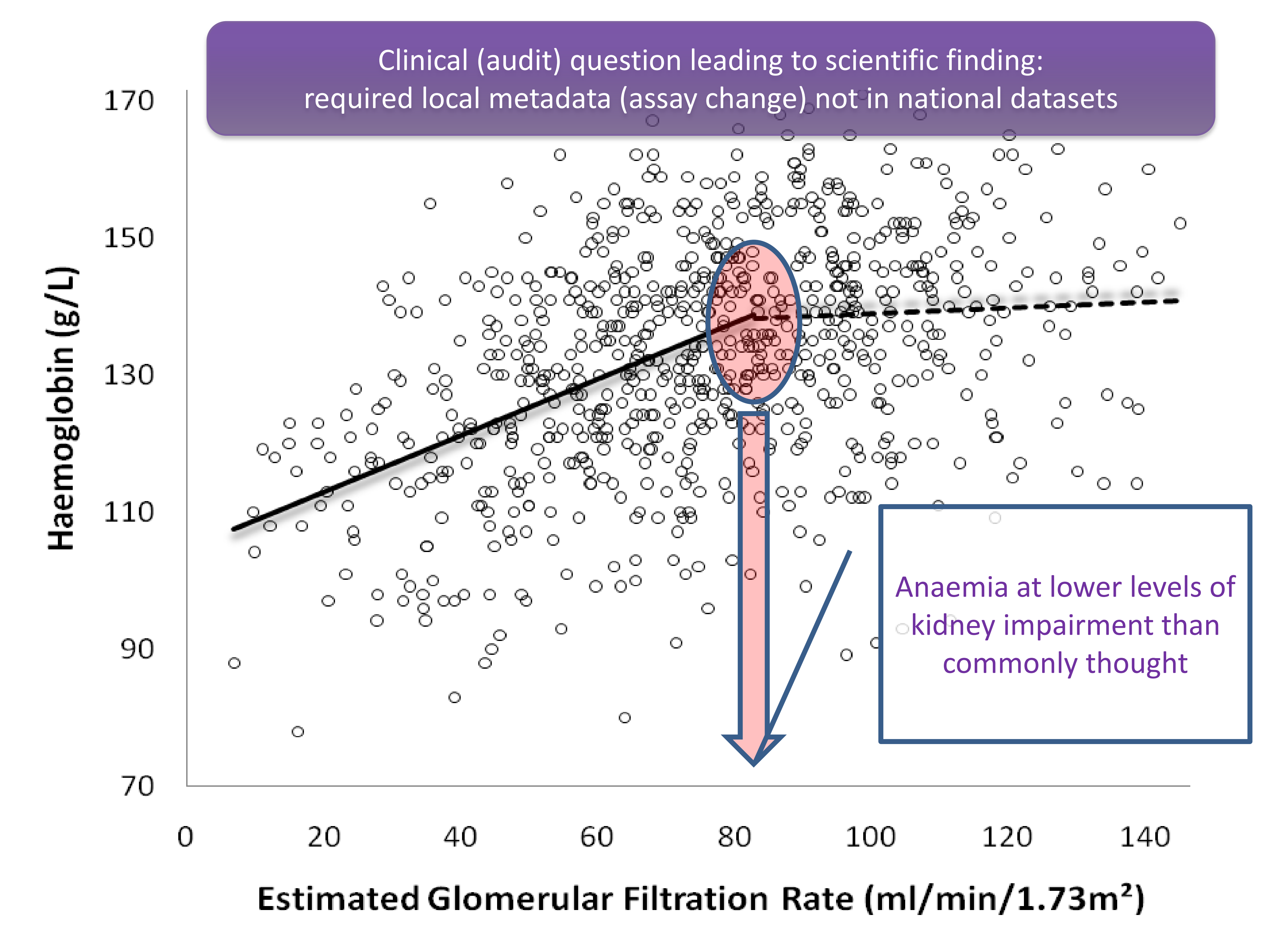
Open Unified Models for Health Policy



Can models be built from literature?



“Not fully”

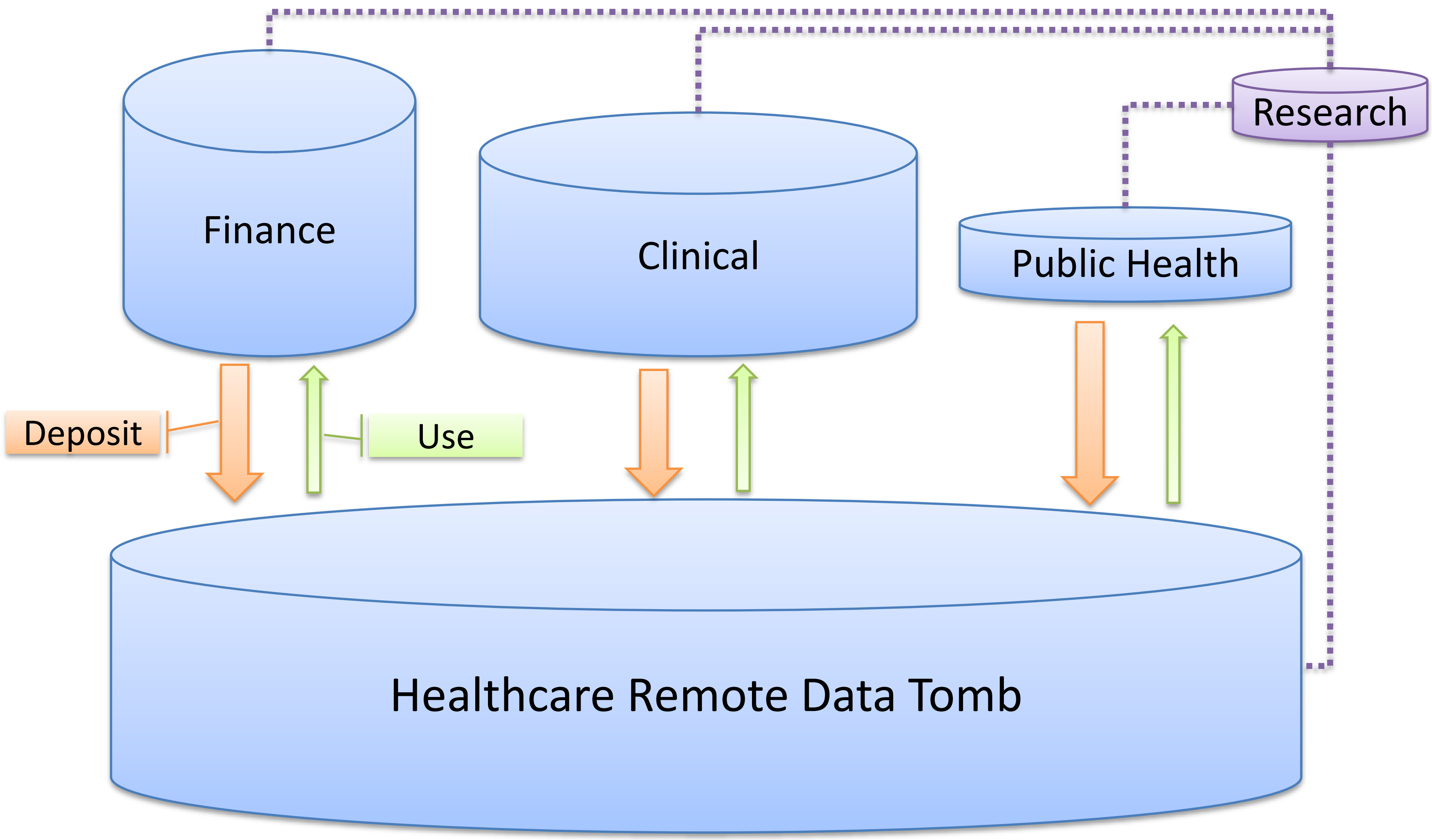


Clinical (audit) question leading to scientific finding:
required local metadata (assay change) not in national datasets
Haemoglobin (g/L)
170
150
130
110
90
70
0
20
40
60
80
100
120
140
Estimated Glomerular Filtration Rate (ml/min/1.73m²)
Anaemia at lower levels of kidney impairment than commonly thought

Framework for Digital Research in Healthcare: I

DIGITAL & OPERATIONAL INFRASTRUCTURE: E-LAB

Digital Dust (data deposit > use)

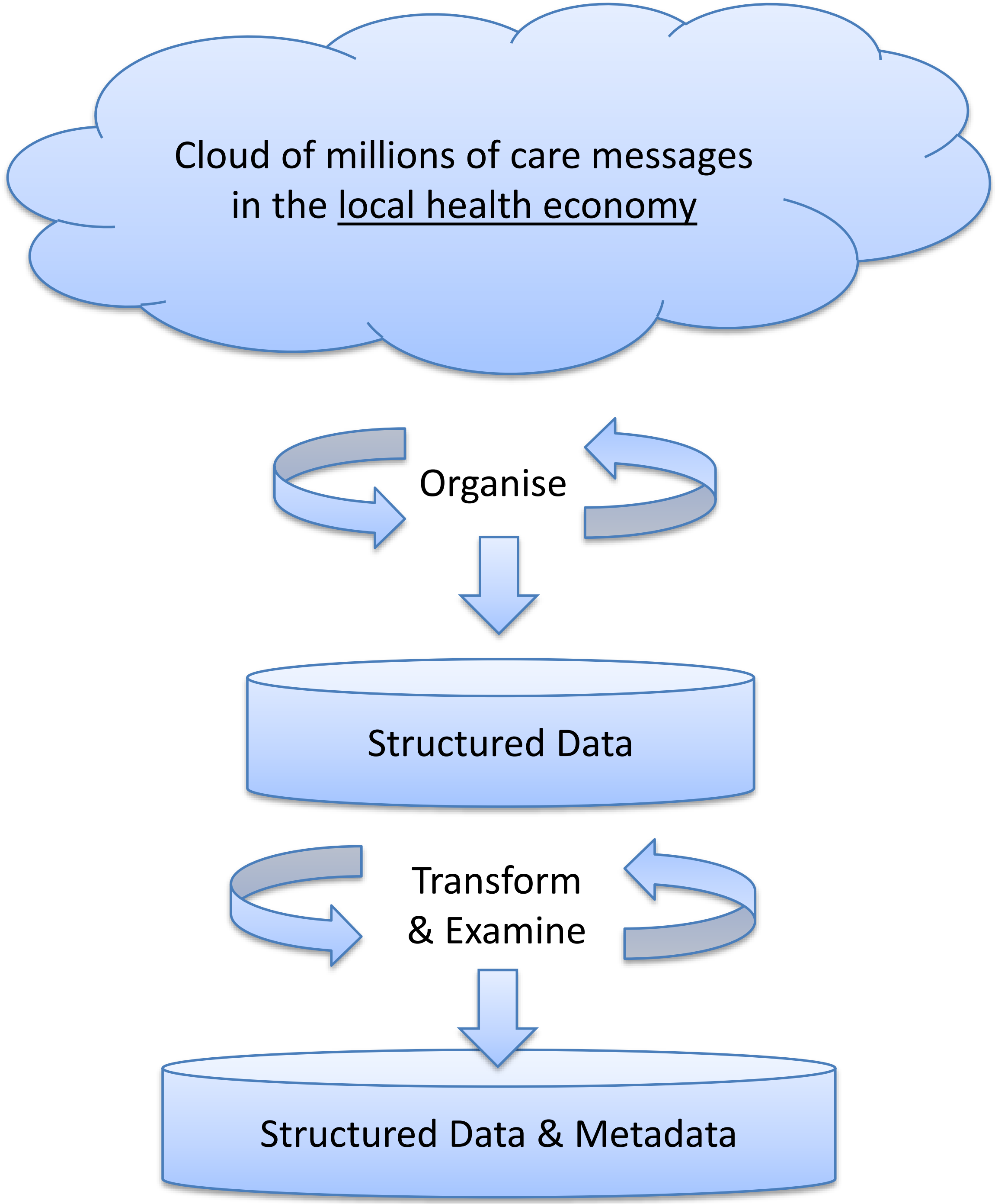


Data supply is not the bottleneck

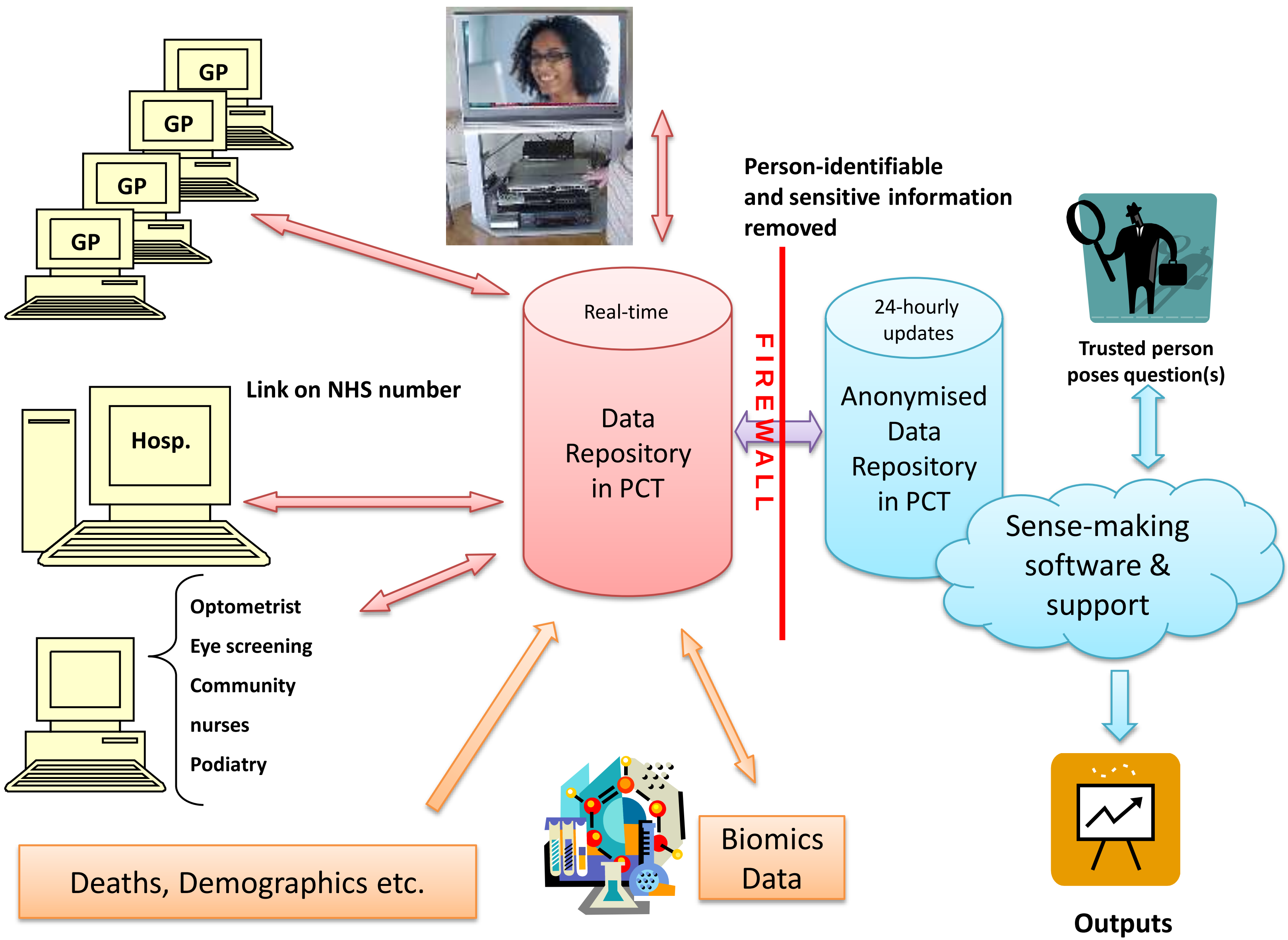
Methods/Models ↑

Contextual expertise ↔

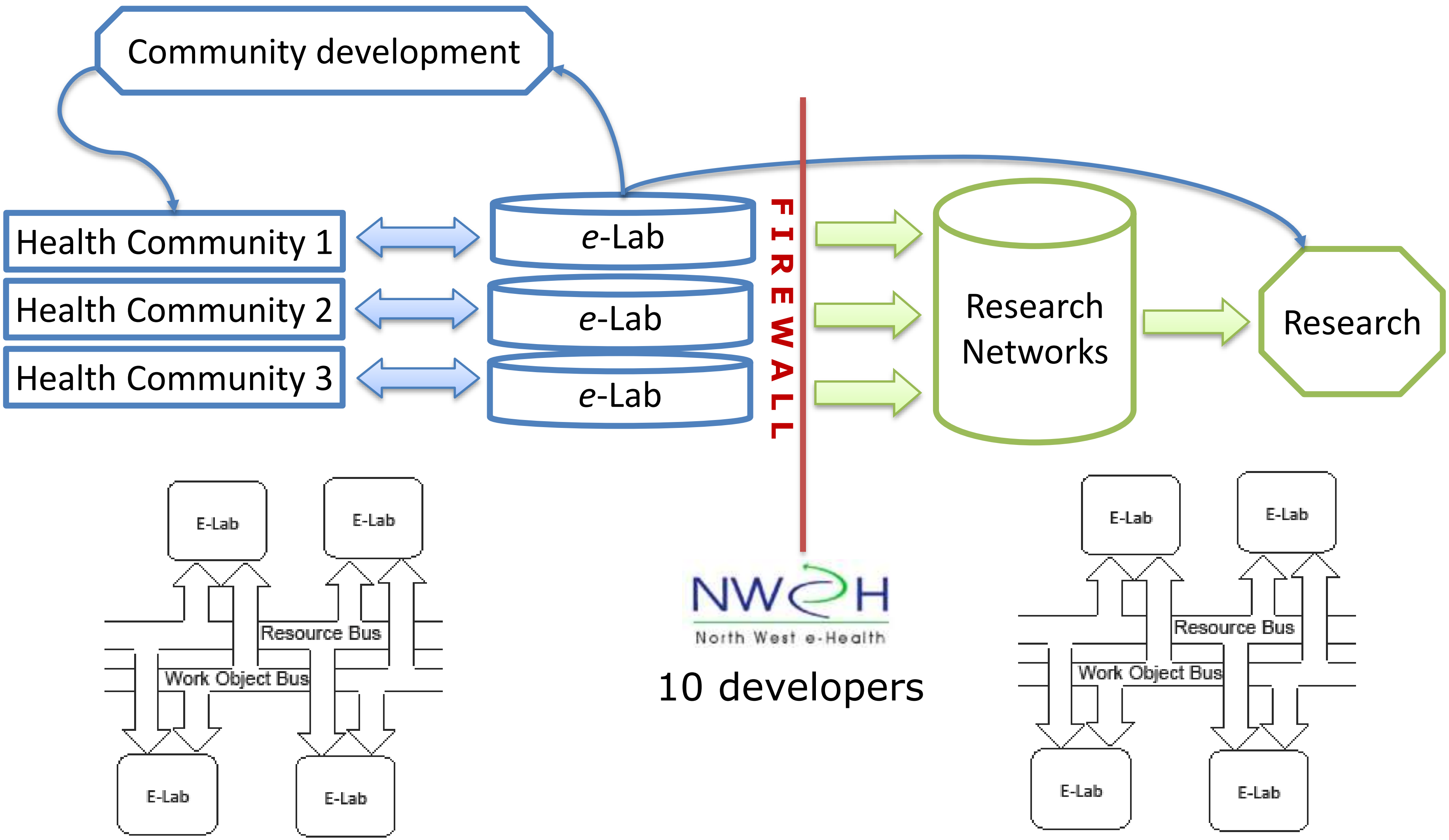
Data ↑↑↑


Cloud of millions of care messages
in the local health economy
Organise
Structured Data
Transform
& Examine
Structured Data & Metadata

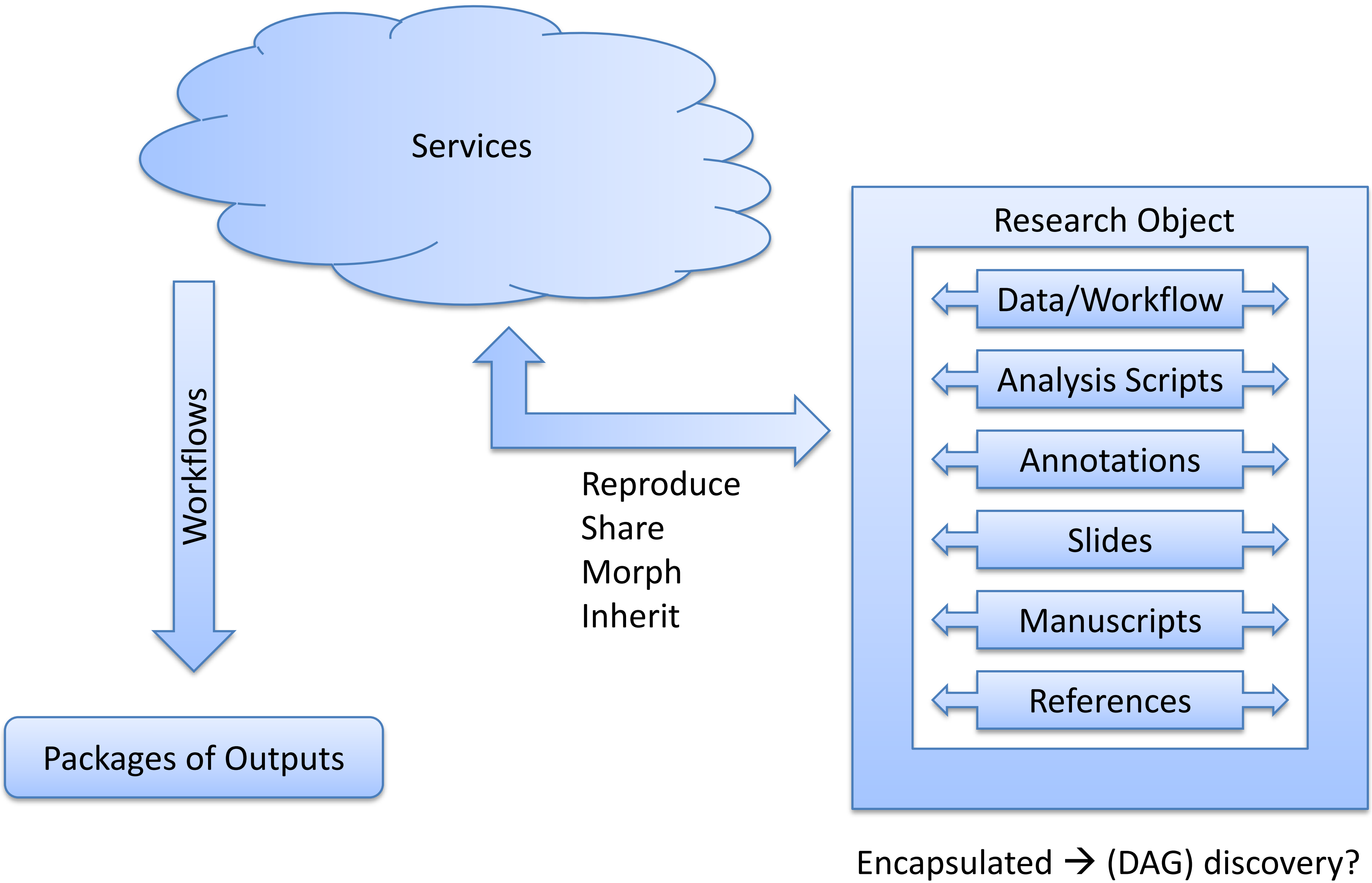
NHS e-Lab: Salford Pilot



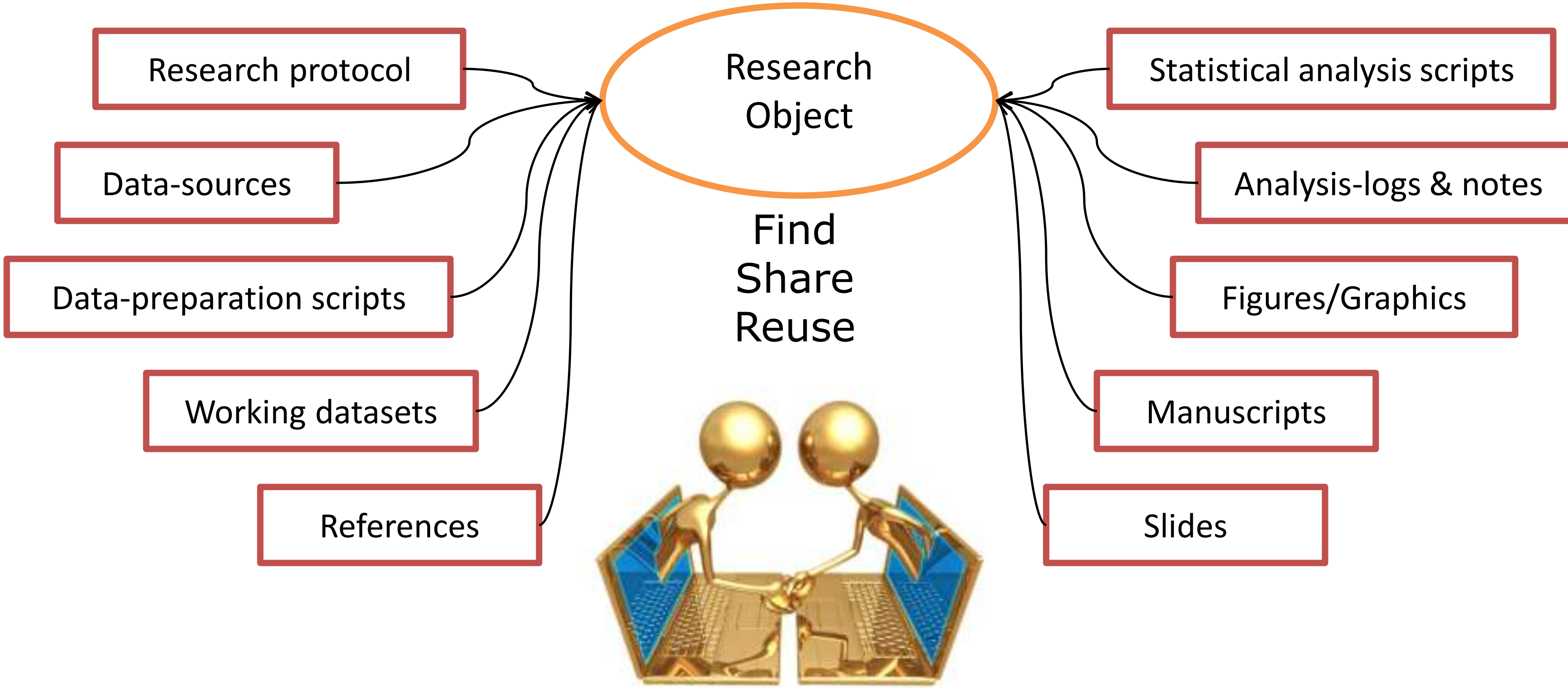
Federation: More local use → better quality data



Work/Research Object



e-Lab



Socially-stimulating science, in-silico

Prototype NHS e-Lab

NW H eLab Life Expectancy In Salford

gary User Admin Profile Log Out

settings

Project Details
- **Data**
- Documents
- Data Exploration
- Notes
- Snapshots
- People

Help

Data

Life Expectancy

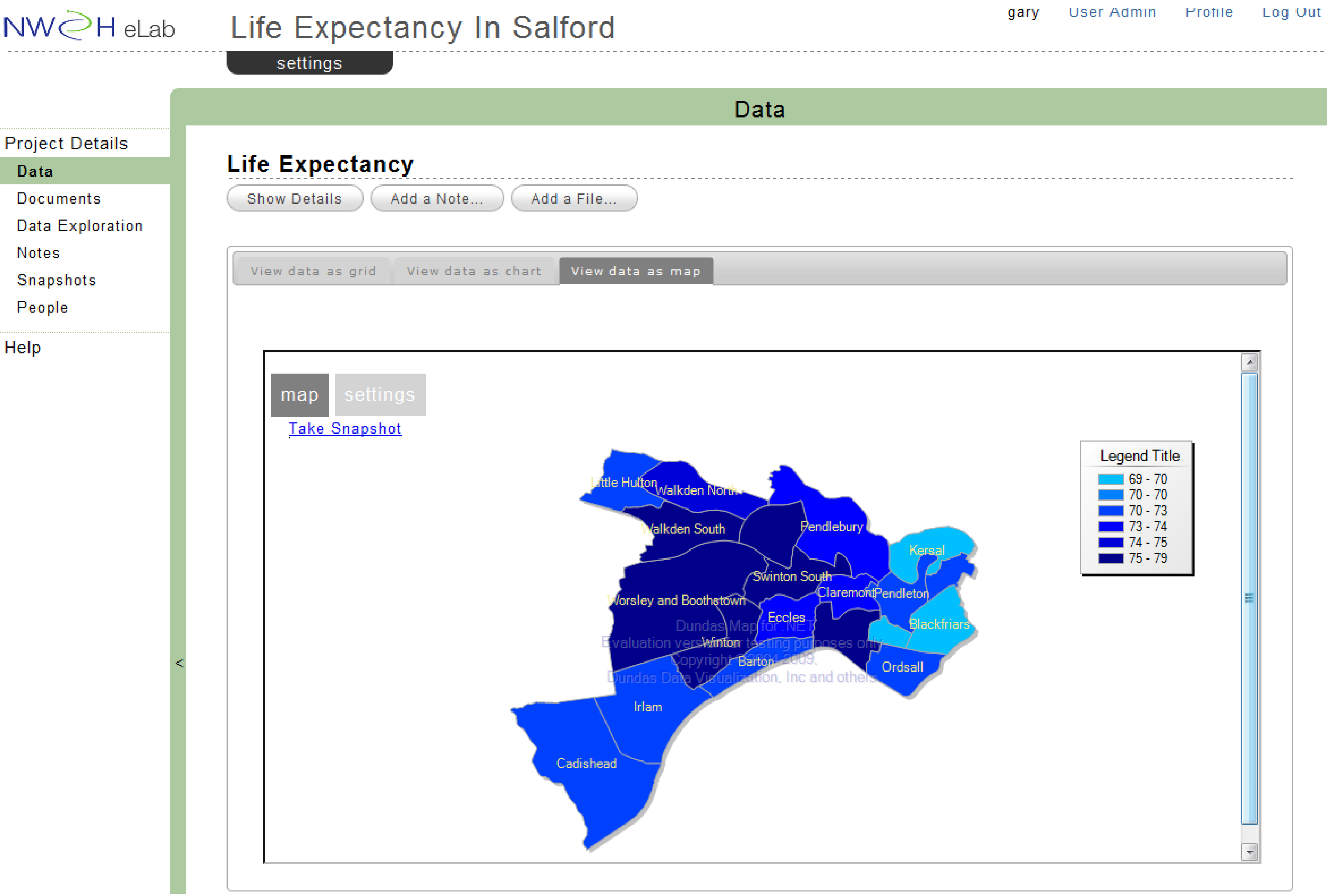
Show Details | Add a Note... | Add a File...

View data as grid | View data as chart | View data as map

WardCode	WardName	LifeExpectanc	DeprivationSc	IsHighDepriva	IsHighDepriva	DiabetesPrev	CvdPrevalenc	DiabetesPrev	DiabetesPrev(
00BRFA	Barton	73	3.342892	false	true	8	10	9	7
00BRFB	Blackfriars	70	7.474104	true	true	5	6	6	4
00BRFC	Broughton	70.1	8.241365	false	true	10	15	11	9
00BRFD	Cadishead	72.3	1.498242	true	false	3	4	5	2
00BRFE	Claremont	74	0.4376208	false	false	3	5	4	2
00BRFF	Eccles	73.4	2.245419	false	false	3	4	5	2
00BRFG	Irlam	70.2	0.8358688	false	false	3	3	4	2

Page 1 of 2 10

View 1 - 10 of 20



Next: Merge with visualisation research

Digital Curators Promoting Healthcare Innovation: I

ETHICS

Ethical Principles

- Respect for autonomy
- Beneficence
- Non-maleficence
- Justice

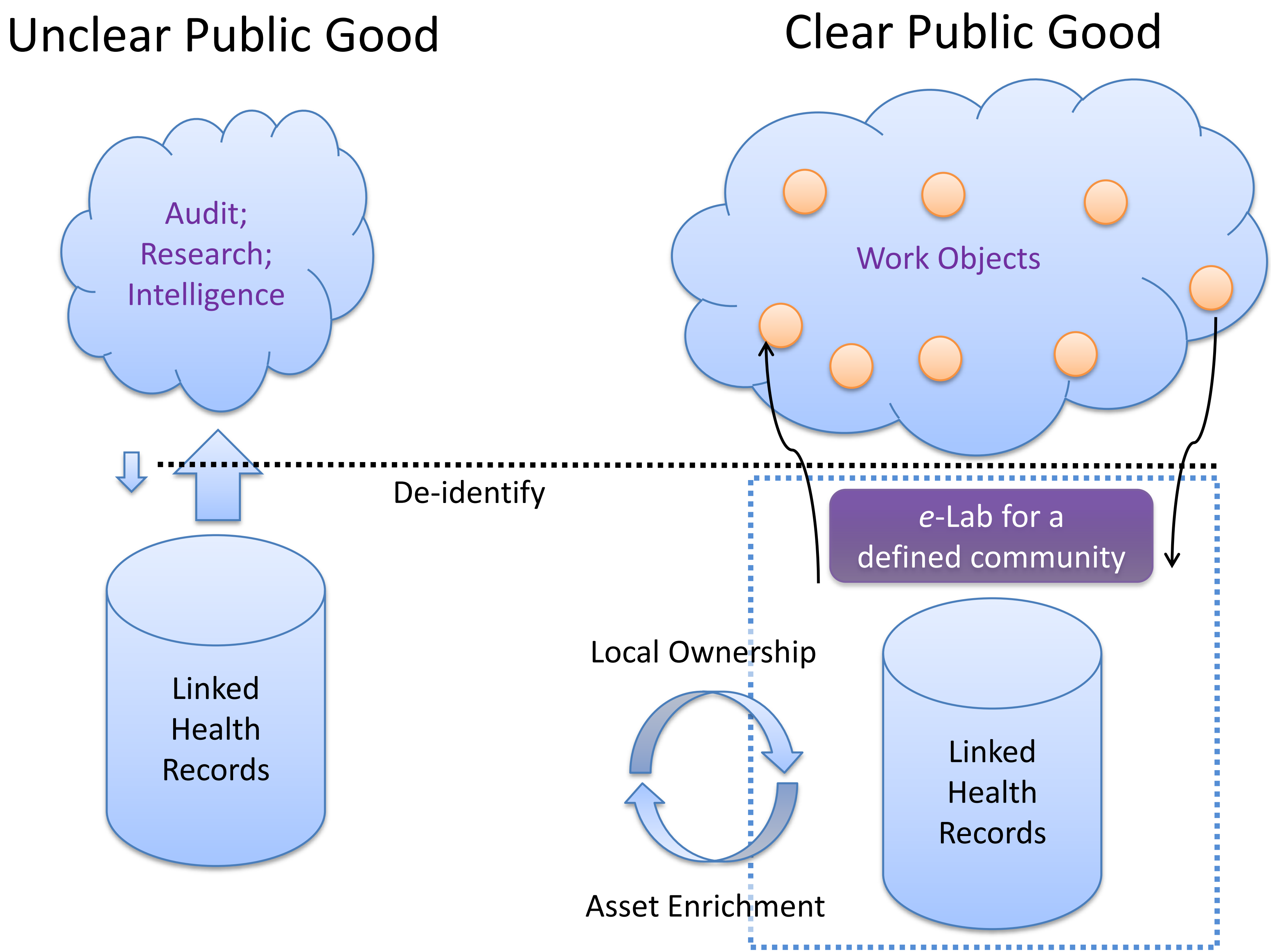
Beauchamp and Childress; Principles Biomedical Ethics, OUP, 5th edition 2001

Respect for Autonomy

- Patient/subject
  - Consent
    - Opt-in
    - Opt-out
  - Right to participate or not
  - Advocates where appropriate
    - Clinical; carer; guardian; data custodian
- Investigator
  - Access to patients/citizens?

Beneficence

- Duty to deliver good for the data donors
- Under-use is unethical
- Audit the context of use
- Measuring good
  - Research quality
  - Clinical utility
  - Patient/citizen involvement



Unclear Public Good
Clear Public Good
Audit;
Research;
Intelligence
Work Objects
De-identify
e-Lab for a
defined community
Linked
Health
Records
Local Ownership
Asset Enrichment
Linked
Health
Records

Non-maleficence

- Part of NHS & University contracts
- Part of clinical and research information governance protocols
- Not dealt with by restricting access to data

Justice

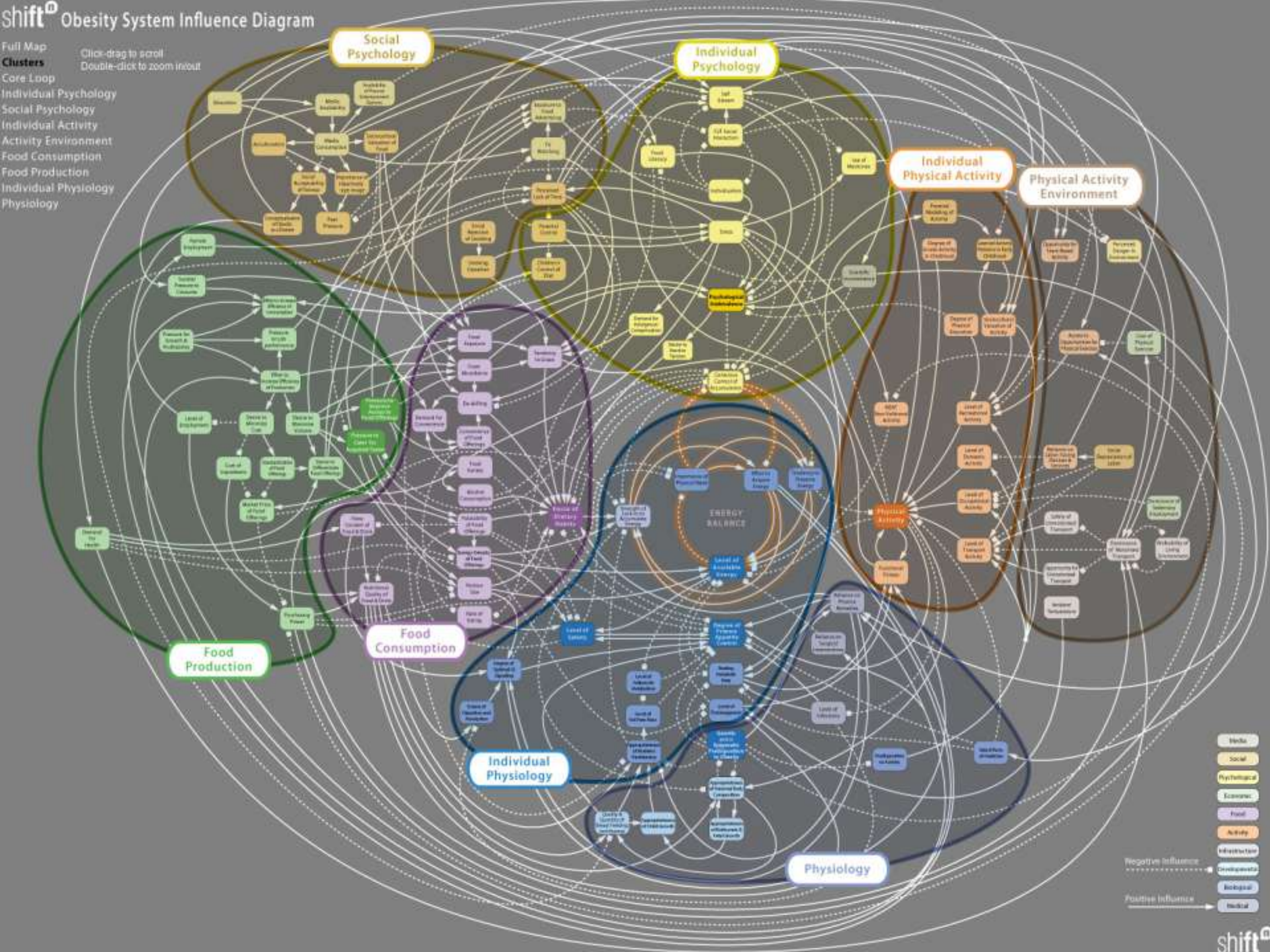
- Knowing the uses of your health records
- Knowing how your practice is measured
- Fair access to data
- Fair access to methods
- Fair access to models
- Intellectual property protection
- Fair networking opportunities for investigators

Data Curation Example: Obesity e-Lab

The e-Science target:

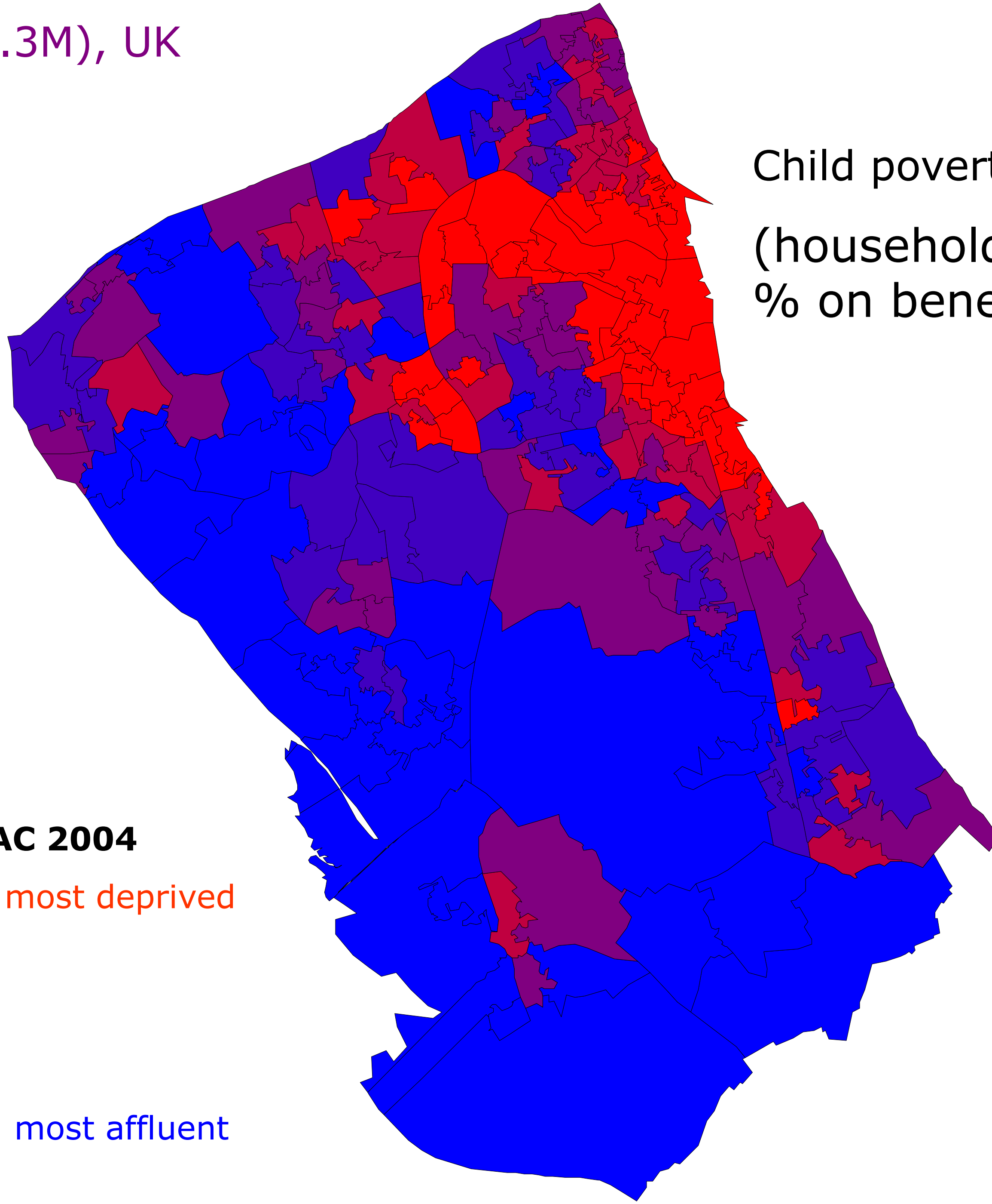
Fragmented understanding of public health problems such as obesity

...data, methods/models and expertise split across disciplines (e.g. Social vs. Biomedical) and settings (e.g. Academia vs. Healthcare)



shift
Obesity System Influence Diagram
Full Map
Clusters
Click-drag to scroll
Double-click to zoom in/out
Core Loop
Individual Psychology
Social Psychology
Individual Activity
Activity Environment
Food Consumption
Food Production
Individual Physiology
Physiology
Social Psychology
Individual Psychology
Individual Physical Activity
Physical Activity Environment
Food Production
Food Consumption
Individual Physiology
Physiology
ENERGY BALANCE
Media
Social
Psychological
Economic
Food
Activity
Infrastructure
Developmental
Biological
Medical
Negative Influence
Positive Influence
shift

Wirral (0.3M), UK



Child poverty map

(households with children: % on benefits in 2001-3)

Fifths of IDAC 2004

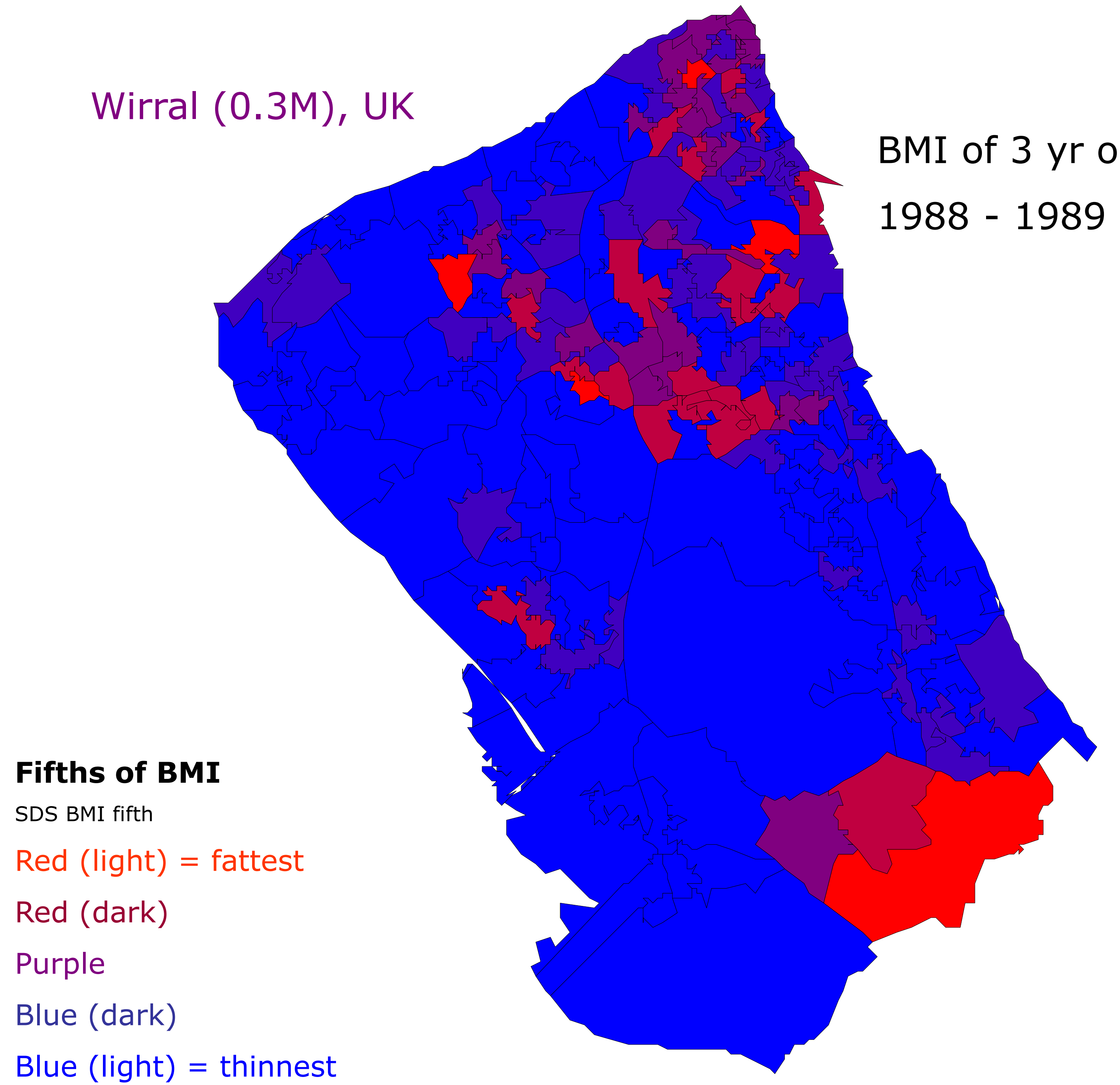
Red (light) = most deprived

Red (dark)

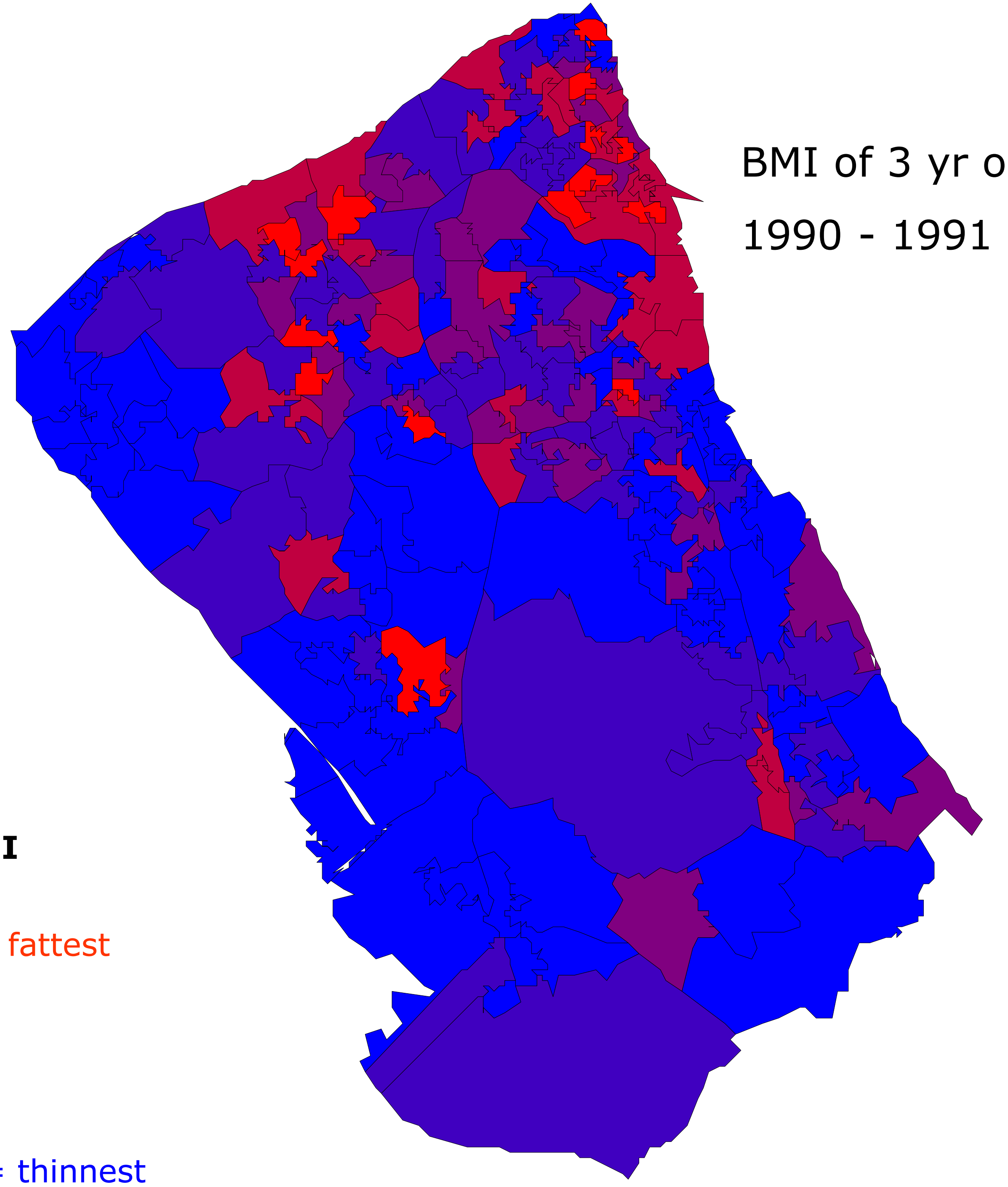
Purple

Blue (dark)

Blue (light) = most affluent



Wirral (0.3M), UK
BMI of 3 yr olds
1988 - 1989
Fifths of BMI
SDS BMI fifth
Red (light) = fattest
Red (dark)
Purple
Blue (dark)
Blue (light) = thinnest



BMI of 3 yr olds

1990 - 1991

Fifths of BMI

SDS BMI fifth

Red (light) = fattest

Red (dark)

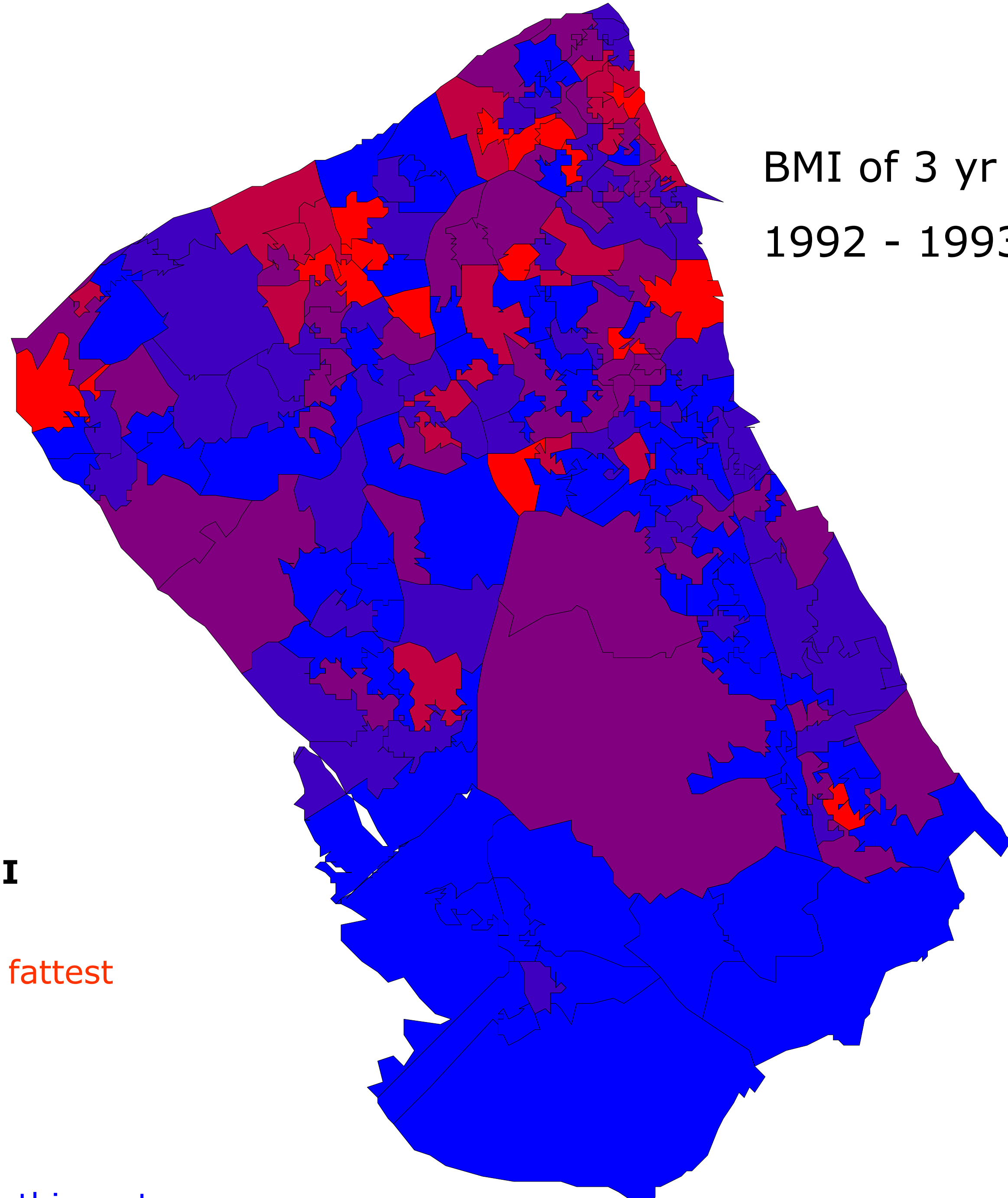
Purple

Blue (dark)

Blue (light) = thinnest

BMI of 3 yr olds

1992 - 1993



Fifths of BMI

SDS BMI fifth

Red (light) = fattest

Red (dark)

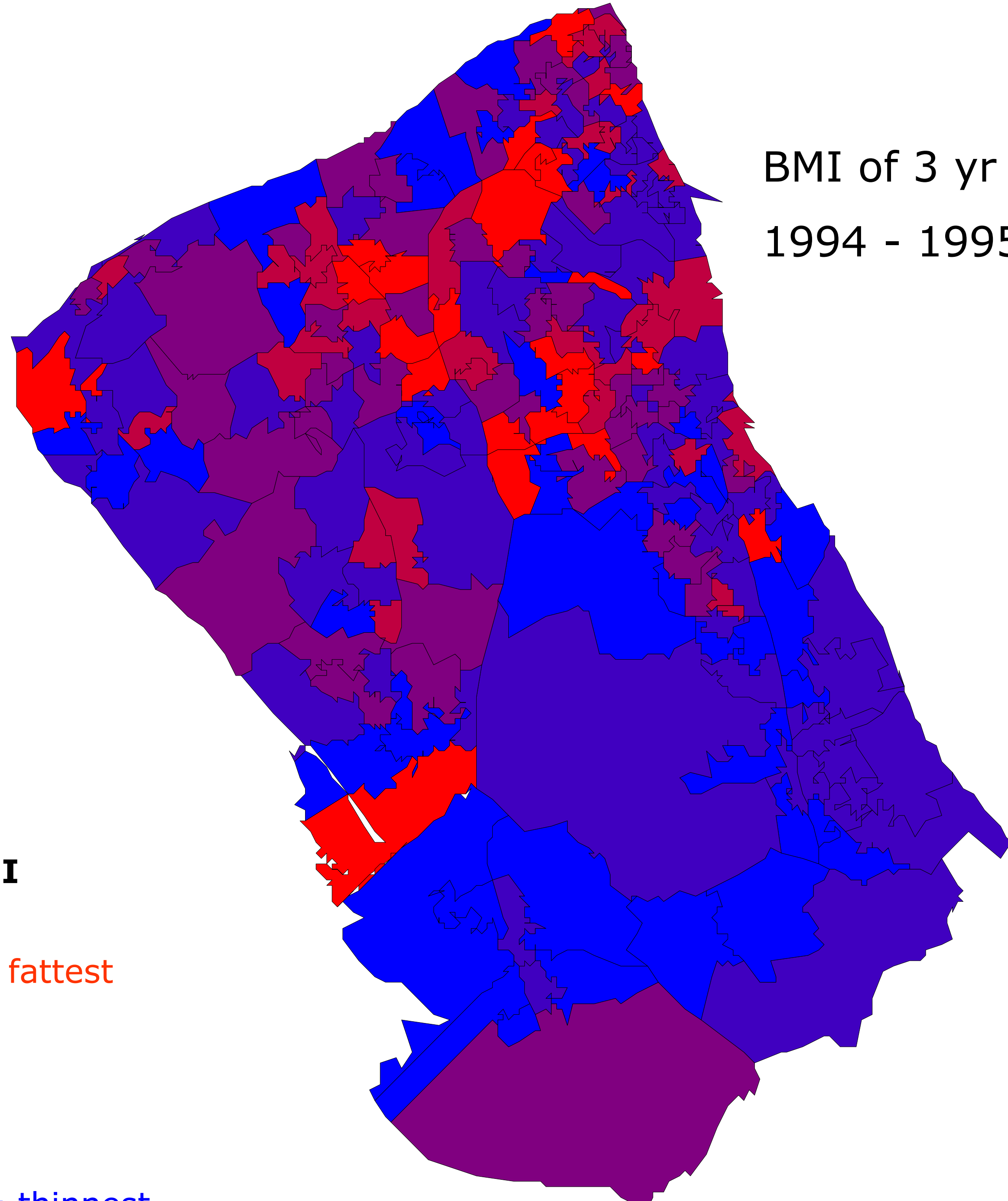
Purple

Blue (dark)

Blue (light) = thinnest

BMI of 3 yr olds

1994 - 1995



Fifths of BMI

SDS BMI fifth

Red (light) = fattest

Red (dark)

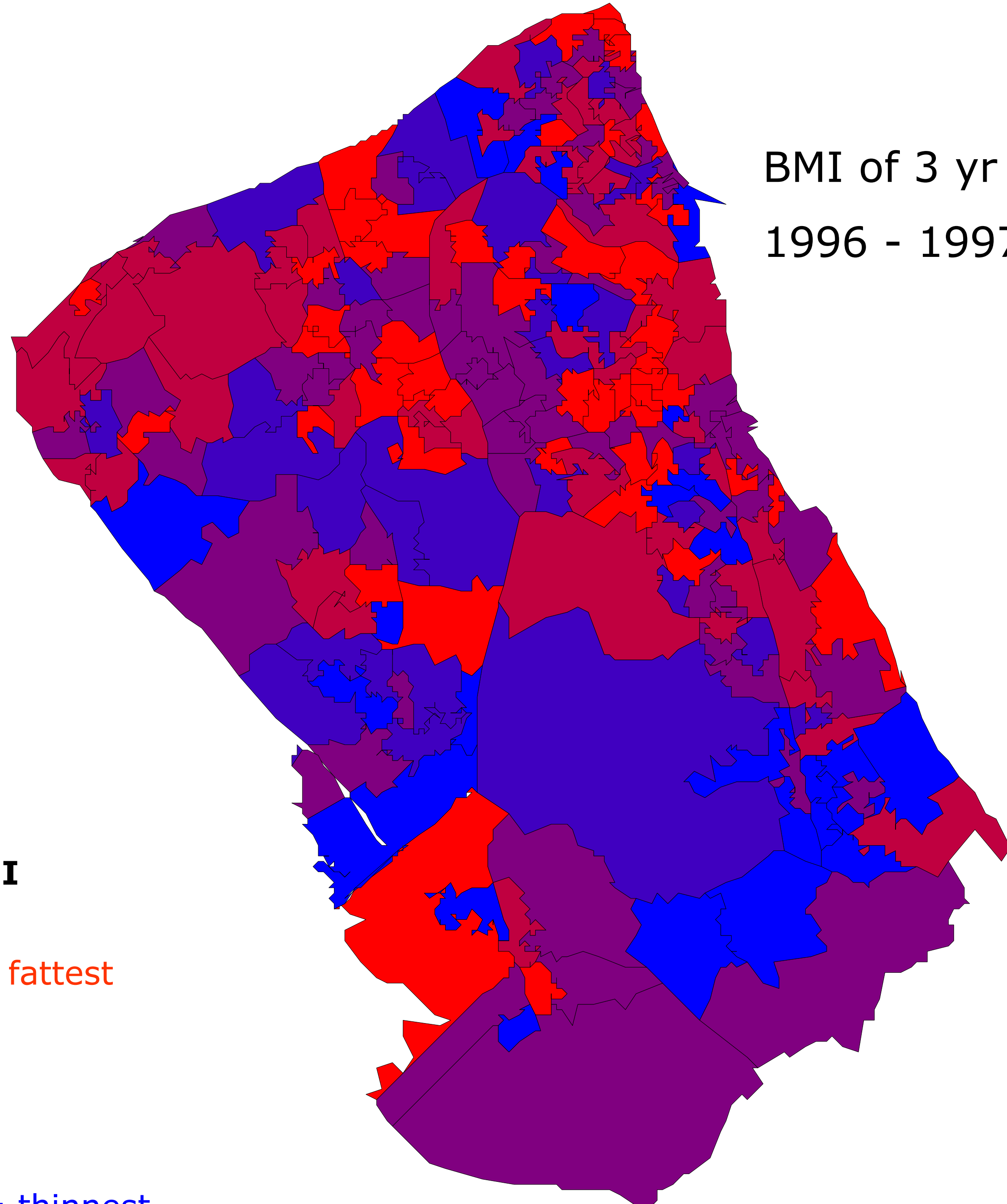
Purple

Blue (dark)

Blue (light) = thinnest

BMI of 3 yr olds

1996 - 1997



Fifths of BMI

SDS BMI fifth

Red (light) = fattest

Red (dark)

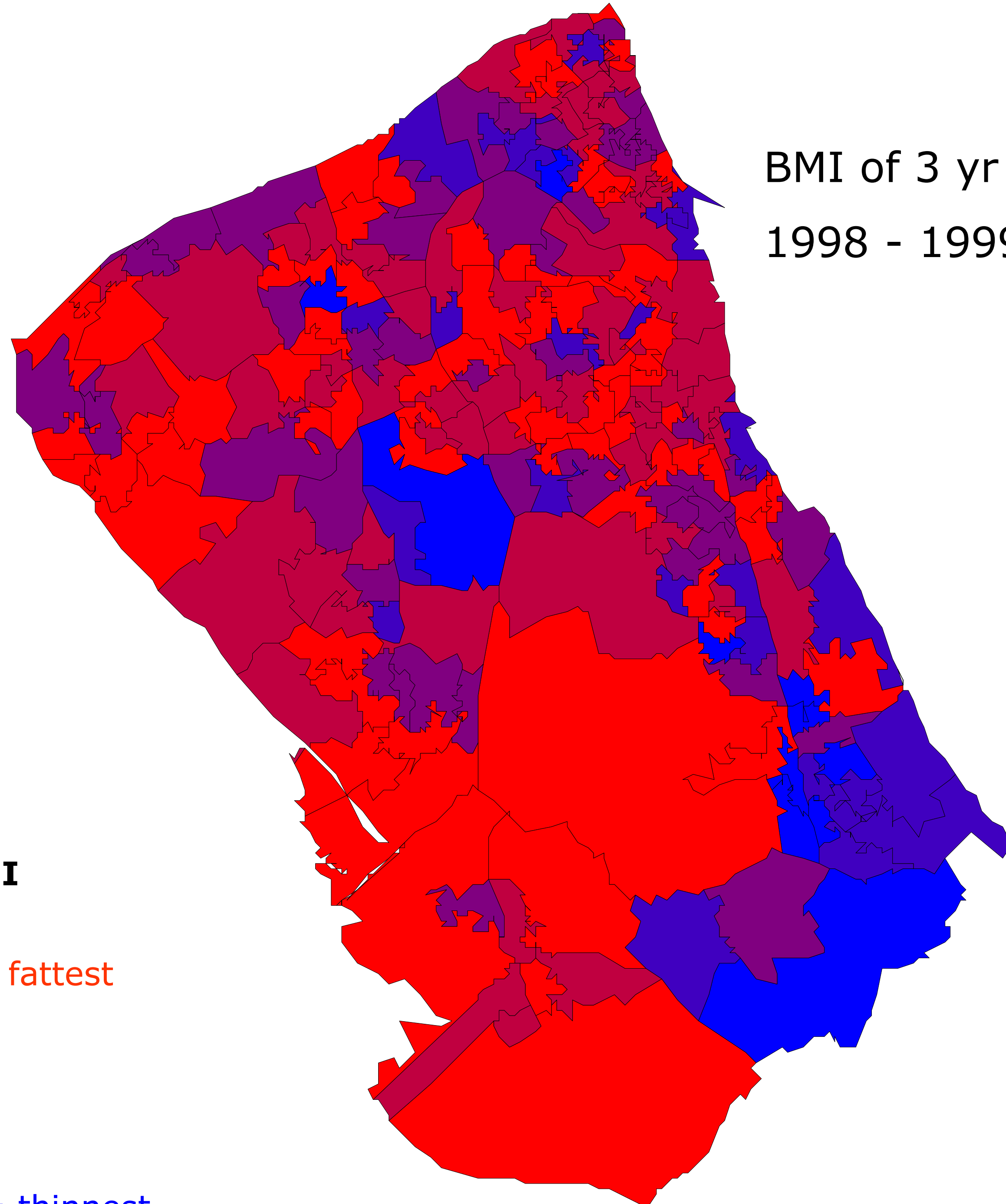
Purple

Blue (dark)

Blue (light) = thinnest

BMI of 3 yr olds

1998 - 1999



Fifths of BMI

SDS BMI fifth

Red (light) = fattest

Red (dark)

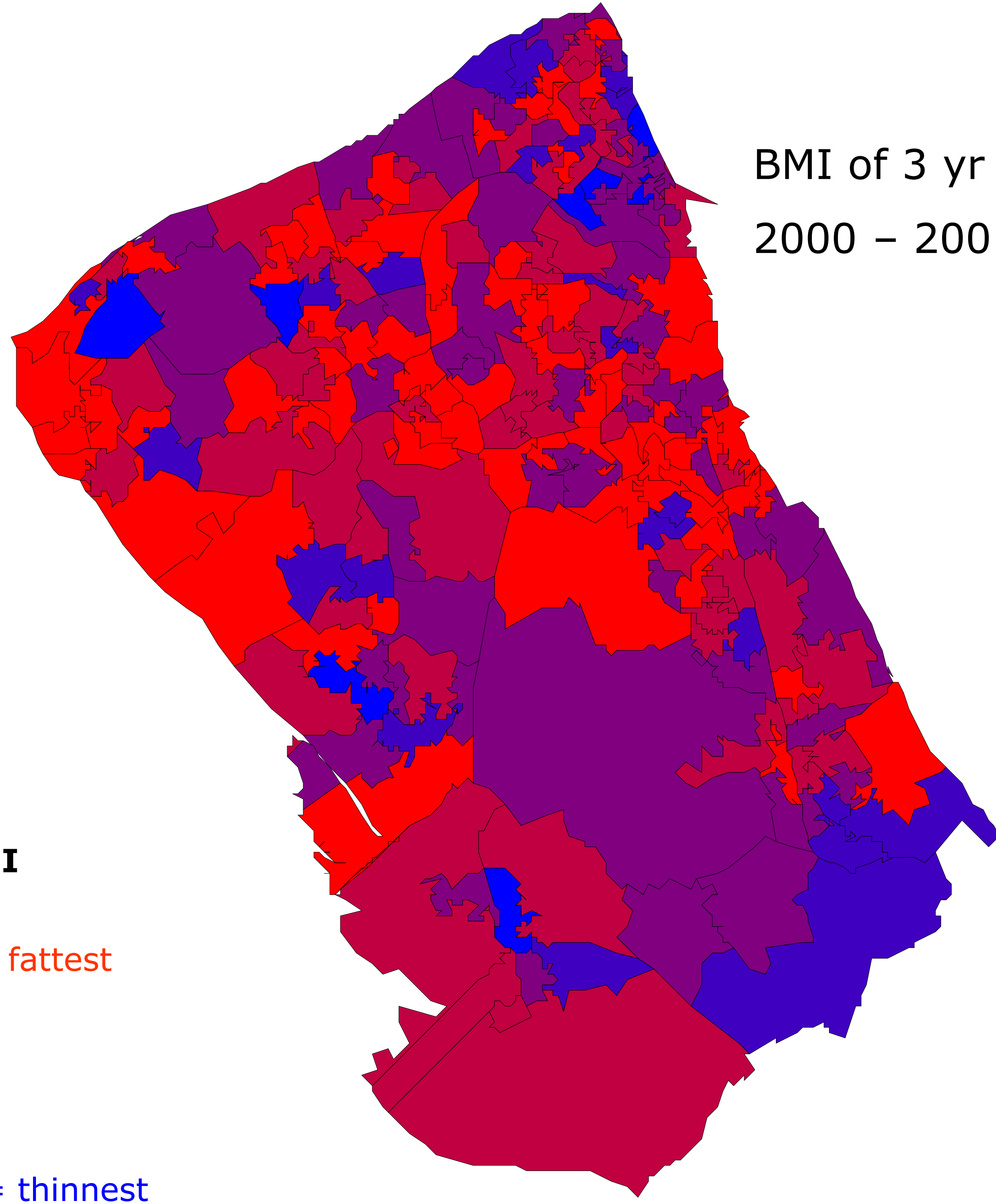
Purple

Blue (dark)

Blue (light) = thinnest

BMI of 3 yr olds

2000 – 2001



Fifths of BMI

SDS BMI fifth

Red (light) = fattest

Red (dark)

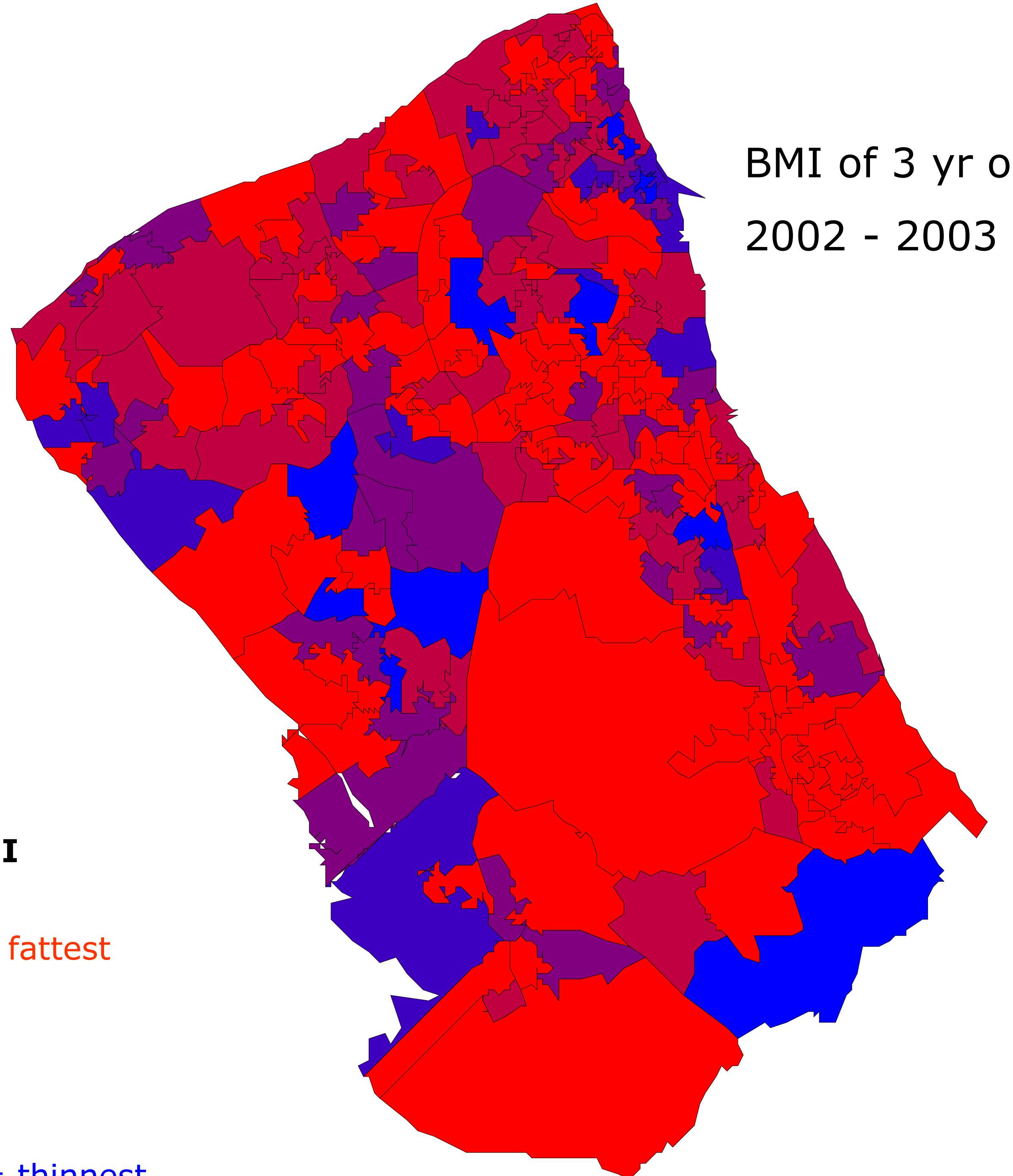
Purple

Blue (dark)

Blue (light) = thinnest

BMI of 3 yr olds

2002 - 2003



Fifths of BMI

SDS BMI fifth

Red (light) = fattest

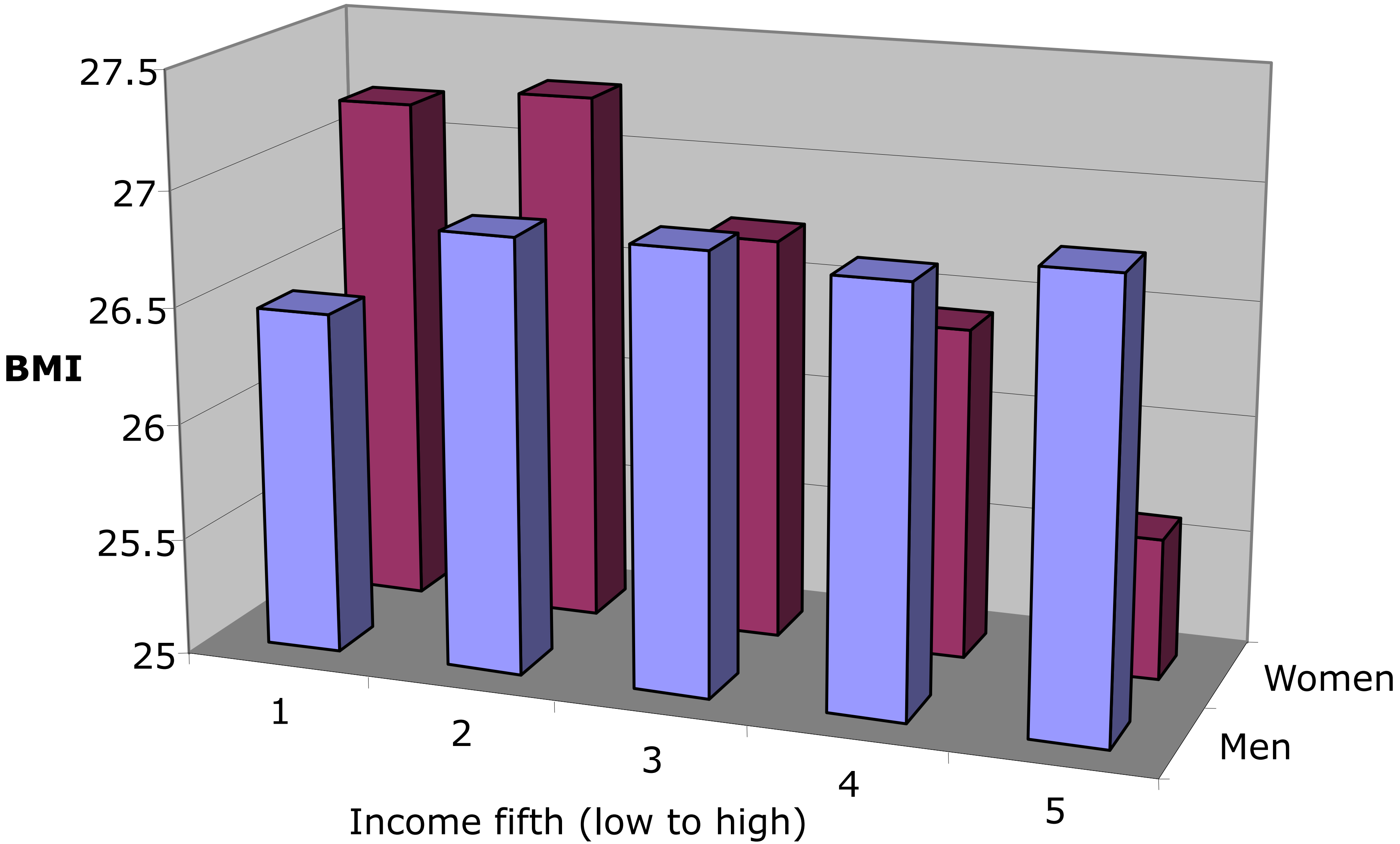
Red (dark)

Purple

Blue (dark)

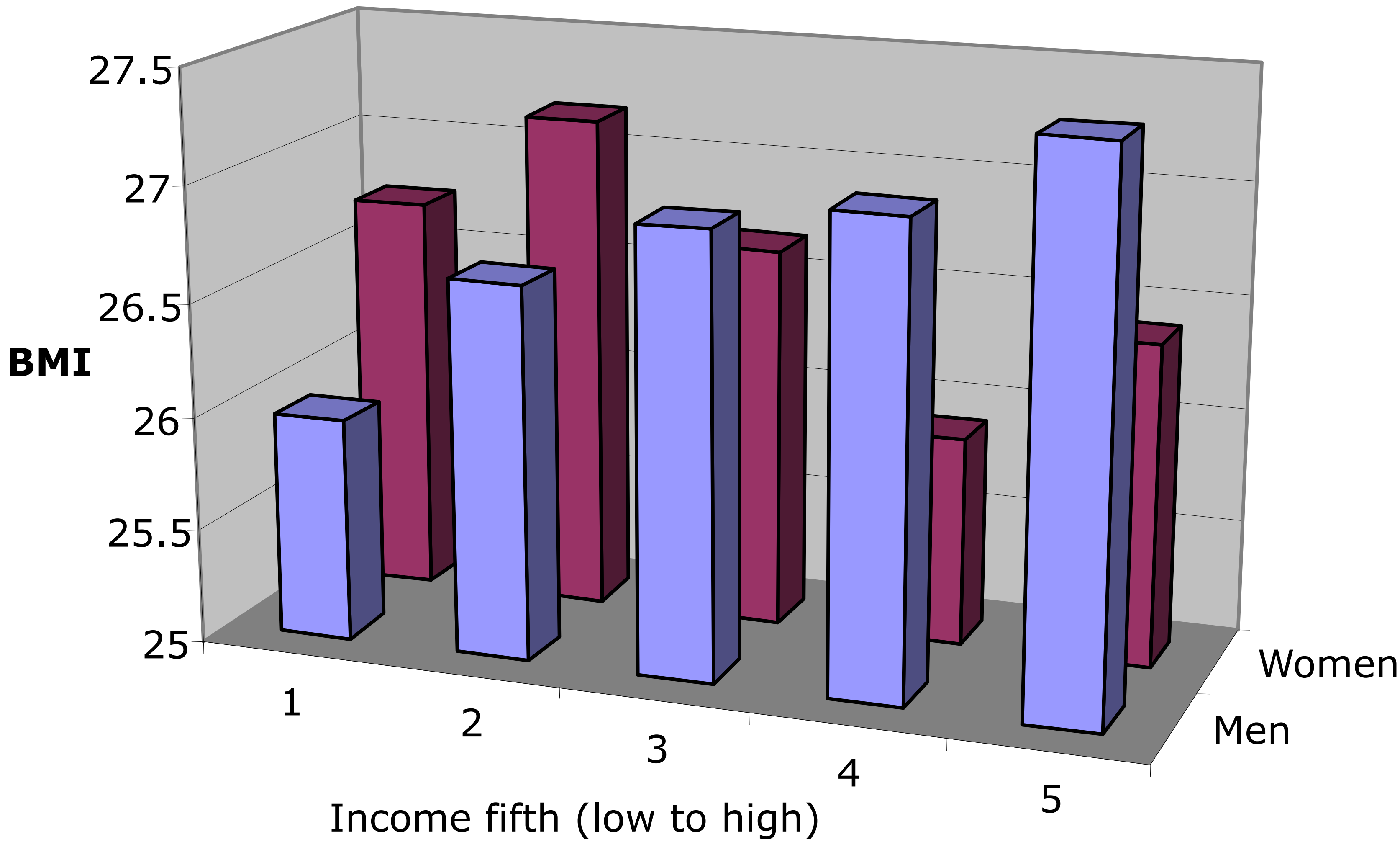
Blue (light) = thinnest

Women and not men from low-income households are fatter in England



Data from Health Survey for England

Women from low-income households and men from high-income households are fatter in Greater Manchester



Data from Health Survey for England

Beware Discipline Clouds

Social Research:
Data, methods & people

Previous slides show
social-biomedical signals
about obesity
from under-used datasets

Biomedical Research:
Data, methods & people

Obesity e-Lab Aim

..to increase the sharing and reuse of

data sources & extracts

and data processing methods

in one in-silico environment (‘e-Lab’)

shared by social and health researchers

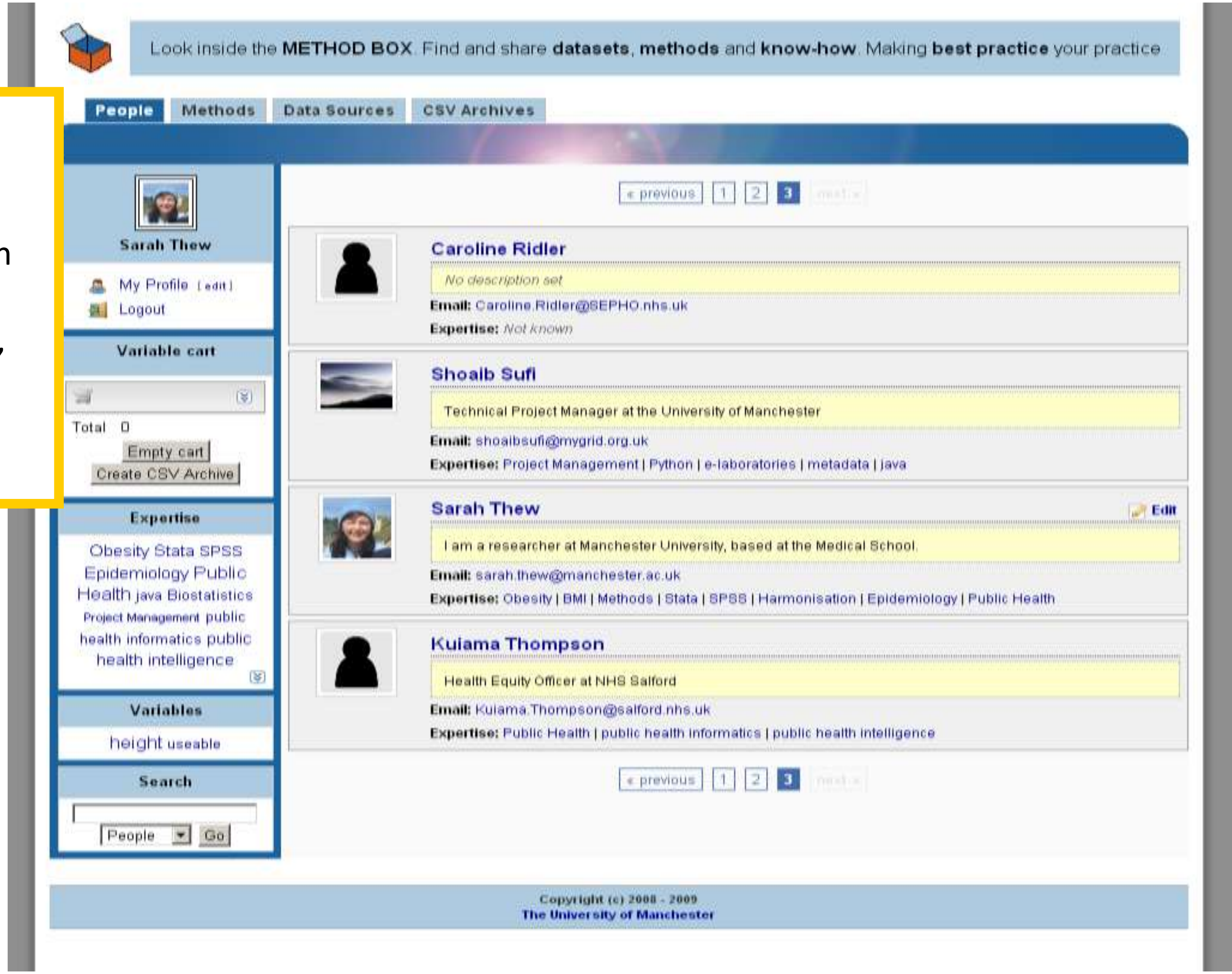
Focus

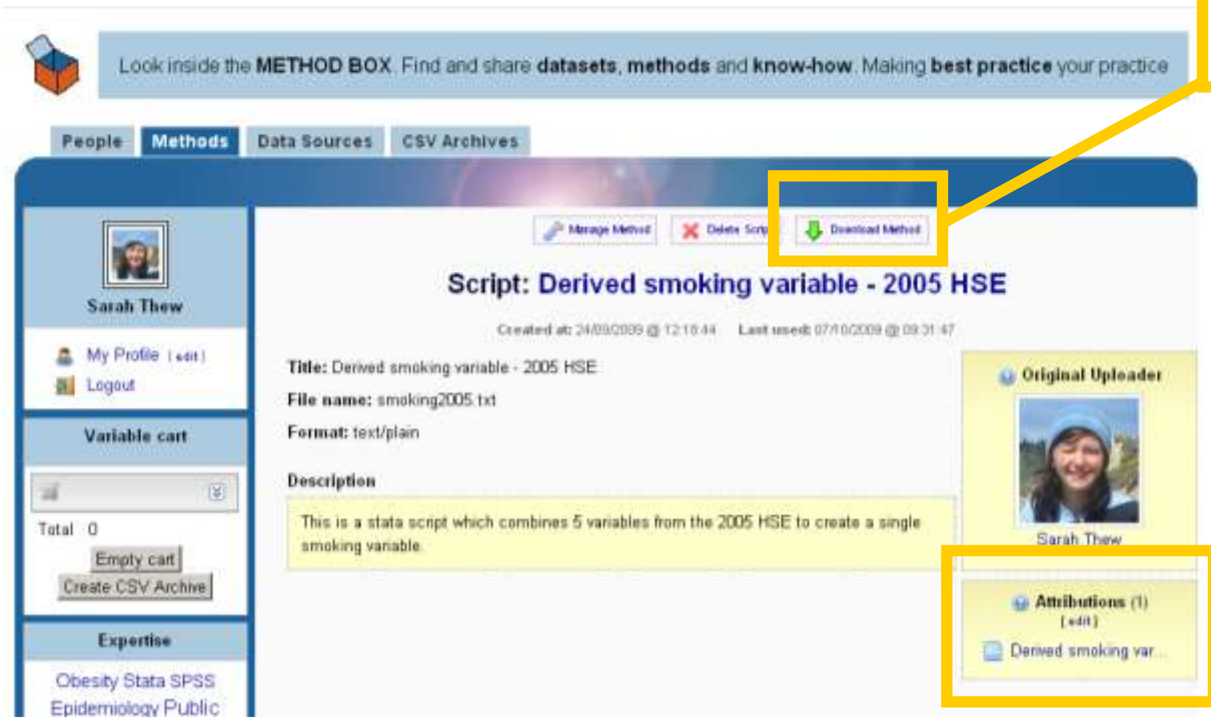
- Health Surveys for England
  - Large-scale (participants * variables)
  - Annual since early 90s
  - Under-used by NHS who fund it
  - Key barrier: extracting a research-ready subset of data
  - Data archive → playground = e-Lab

Supporting and developing interdisciplinary understanding

Sharing resources – tools, methods, data	First step - sharing of resources
Sharing expertise – discussions and reuse around shared resources	Shared resources provide the basis for discussion
Developing interdisciplinary understanding – language, tacit assumptions, methods	Discussions lead to deeper interdisciplinary understanding
Promoting interdisciplinary working	Understanding of other domains promotes more effective interdisciplinary working

MethodBox users include NHS Public Health analysts and Department of Health Public Health Observatory analysts, as well as academic social scientists and epidemiologists.





Download methods and scripts shared by other Methodbox users

The owner of the script can create links back to other methods and scripts used in the development of this script.

Digital Curators Promoting Healthcare Innovation: II

SECURITY

Major Issues with Clinical Studies

- Bias & generalisability
- >50% over-run
- >30% don’t hit recruitment targets
- Unrealistic feasibility assessment
- Consent-management confusion

openCDMS

An Open Source System for:

- multi-centre remote electronic data collection;
- highly configurable security system employing Role Based Access Control;
- fully customisable data set definition including data elements, validation rules and scheduling;
- fully configurable online randomisation with email and SMS text message notification;
- project management reporting including recruitment, completion and UKCRN accrual;
- on-line and off-line data collection;
- flexible query system for identifying eligible trial participants and nested case-control studies;
- designed for compliance with 21 CFR part 11; EMEA GCP; ISO 27001

openCDMS
clinical data management system

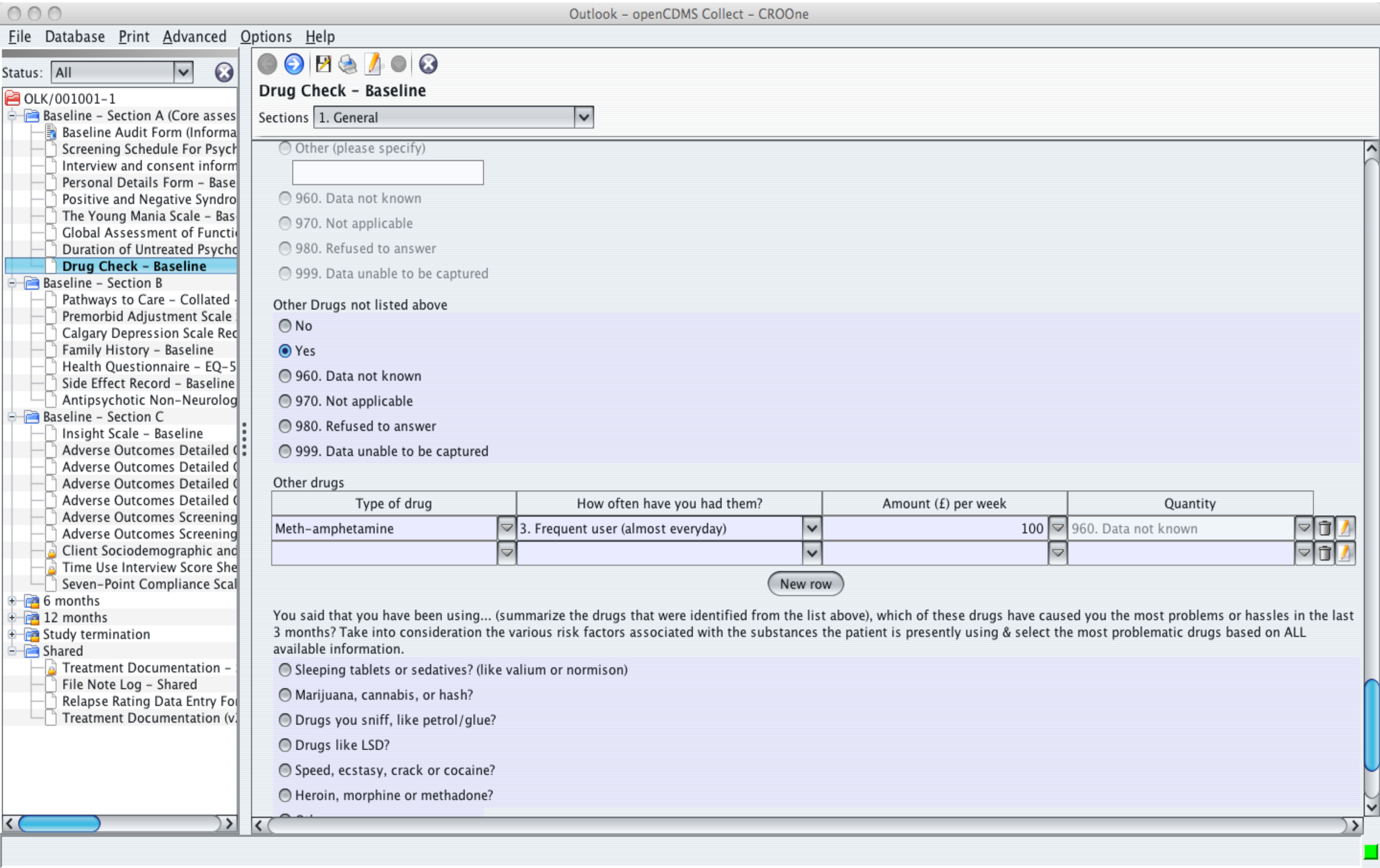
openCDMS in use

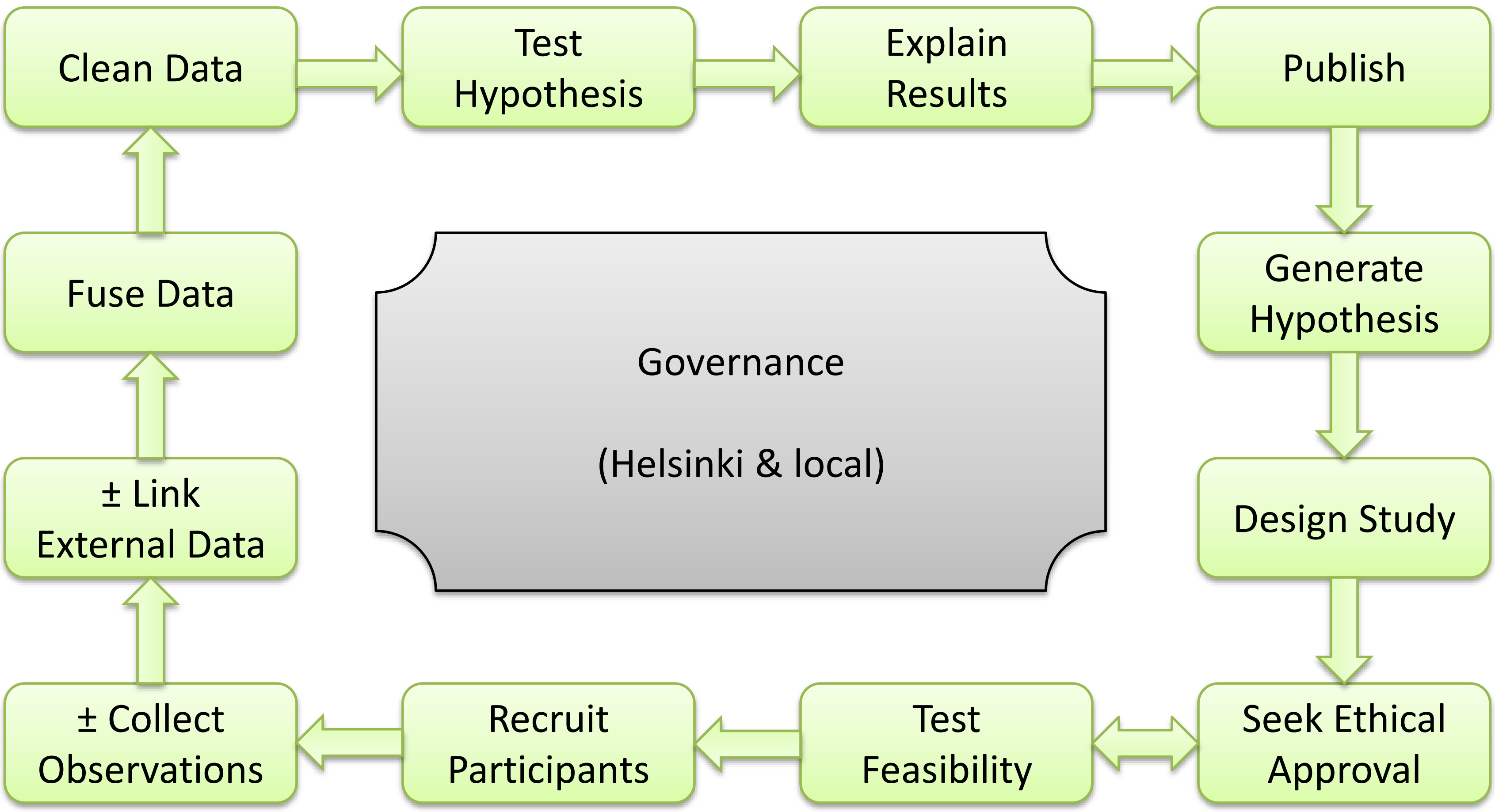
- PsyGrid study – cohort of 700+ schizophrenics followed from first episode for 18 months
- Running numerous mental health trials
- ADDRESS – Type I+II diabetes 10y cohort study
- DARE – Diabetes cohort (phenotype and genotype)

openCDMS
clinical data management system

openCDMS
clinical data management system

Investigator-shaped data capture




Clean Data
Test Hypothesis
Explain Results
Publish
Fuse Data
Governance
(Helsinki & local)
Generate Hypothesis
± Link External Data
Design Study
± Collect Observations
Recruit Participants
Test Feasibility
Seek Ethical Approval

NWeH
North West e-Health

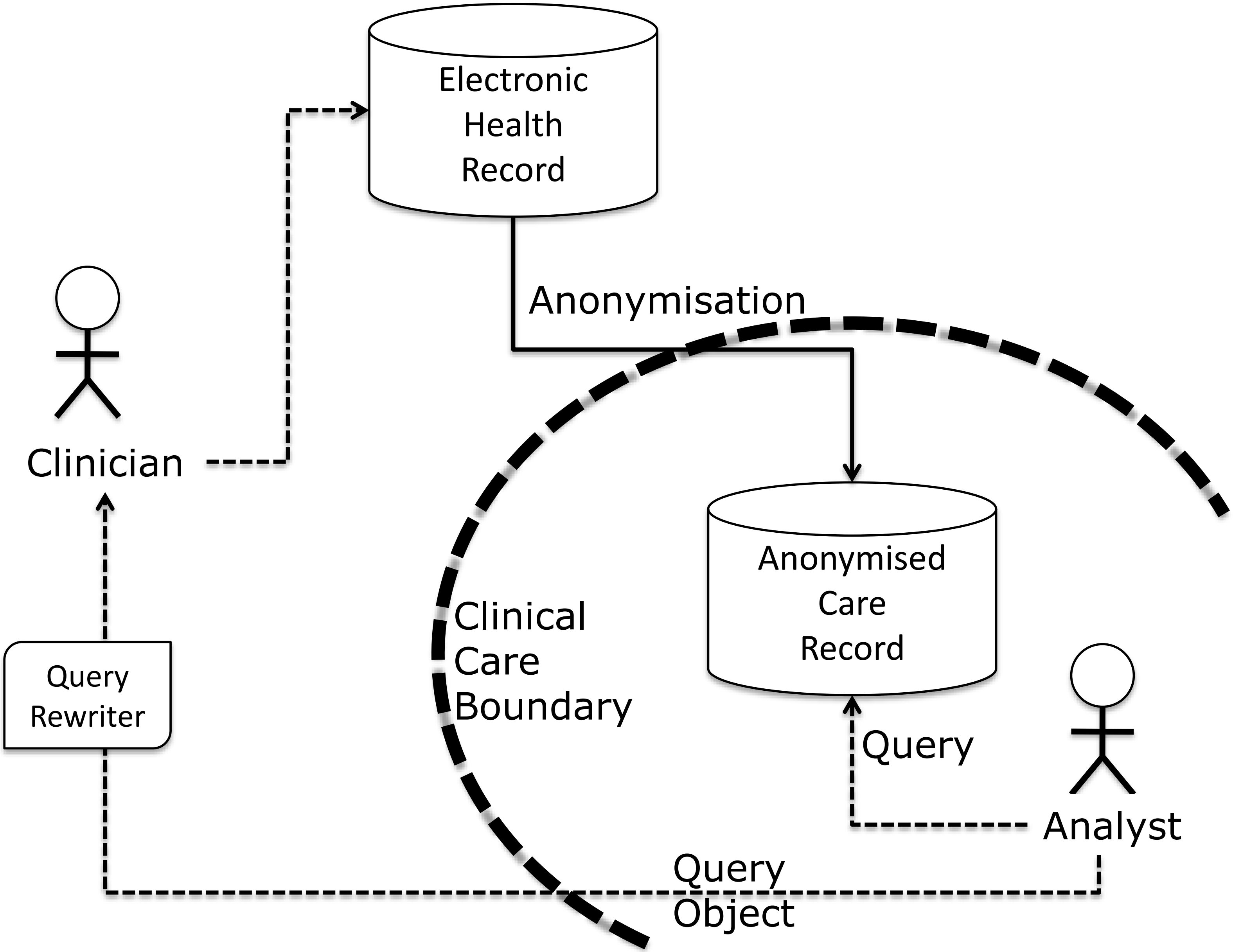
Managing clinical studies

Making sense of local healthcare

openCDMS
clinical data management system

+ community

e-Lab


Electronic
Health
Record
Anonymisation
Clinician
Anonymised
Care
Record
Clinical
Care
Boundary
Query
Rewriter
Query
Analyst
Query
Object

Consent-for-consent

...is the consent to

search an individual's health record

to determine whether or not

they should be invited

to take part in a clinical study.

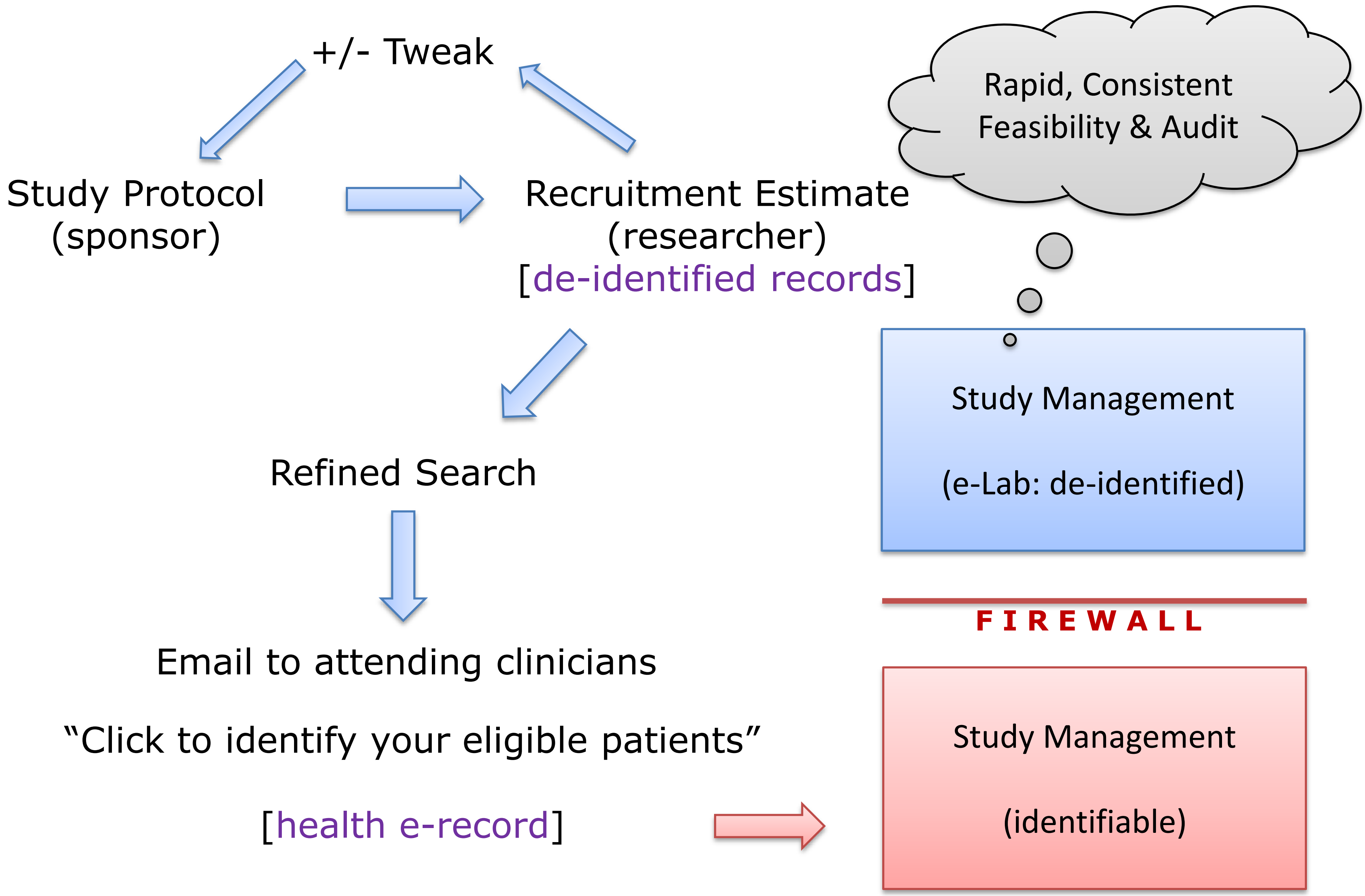
FARSITE

Feasibility **A**ssessment

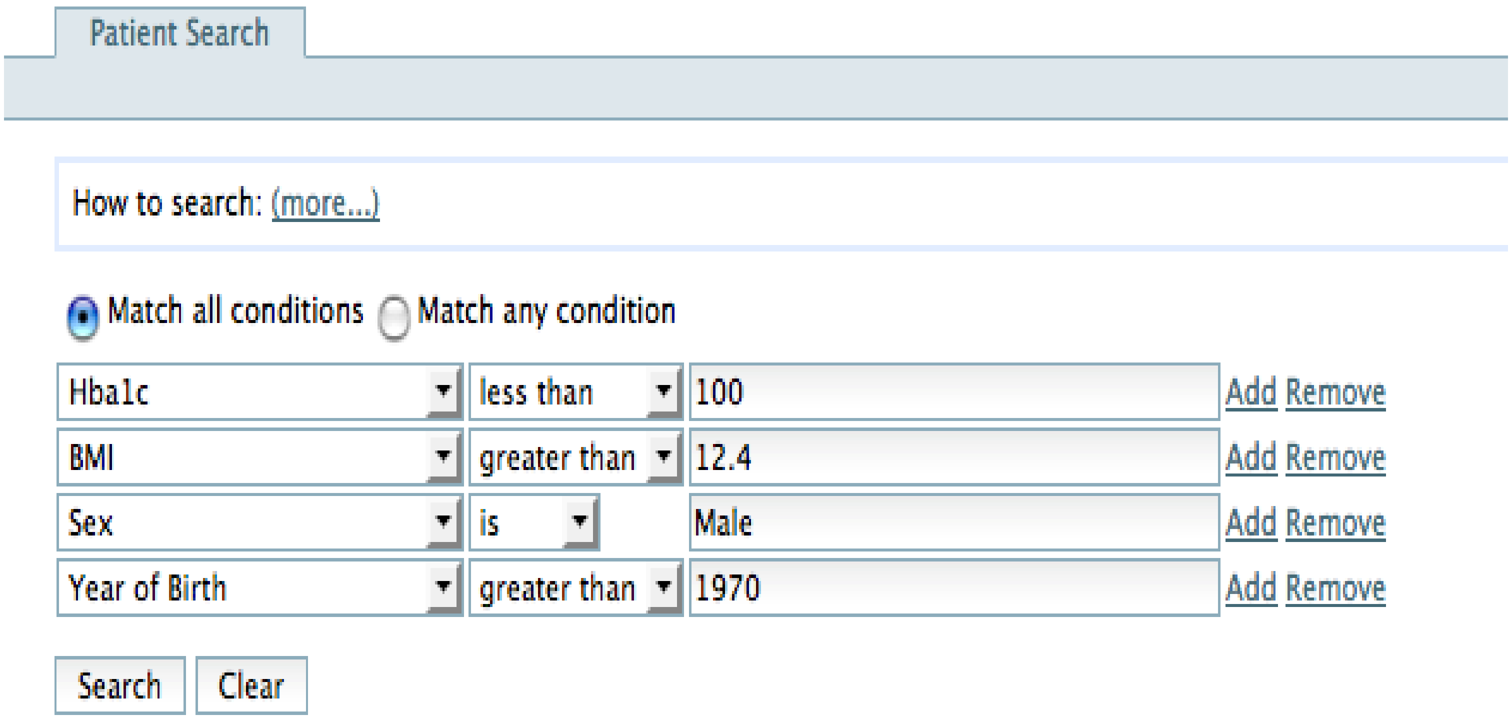
and **R**ecruitment **S**ystem

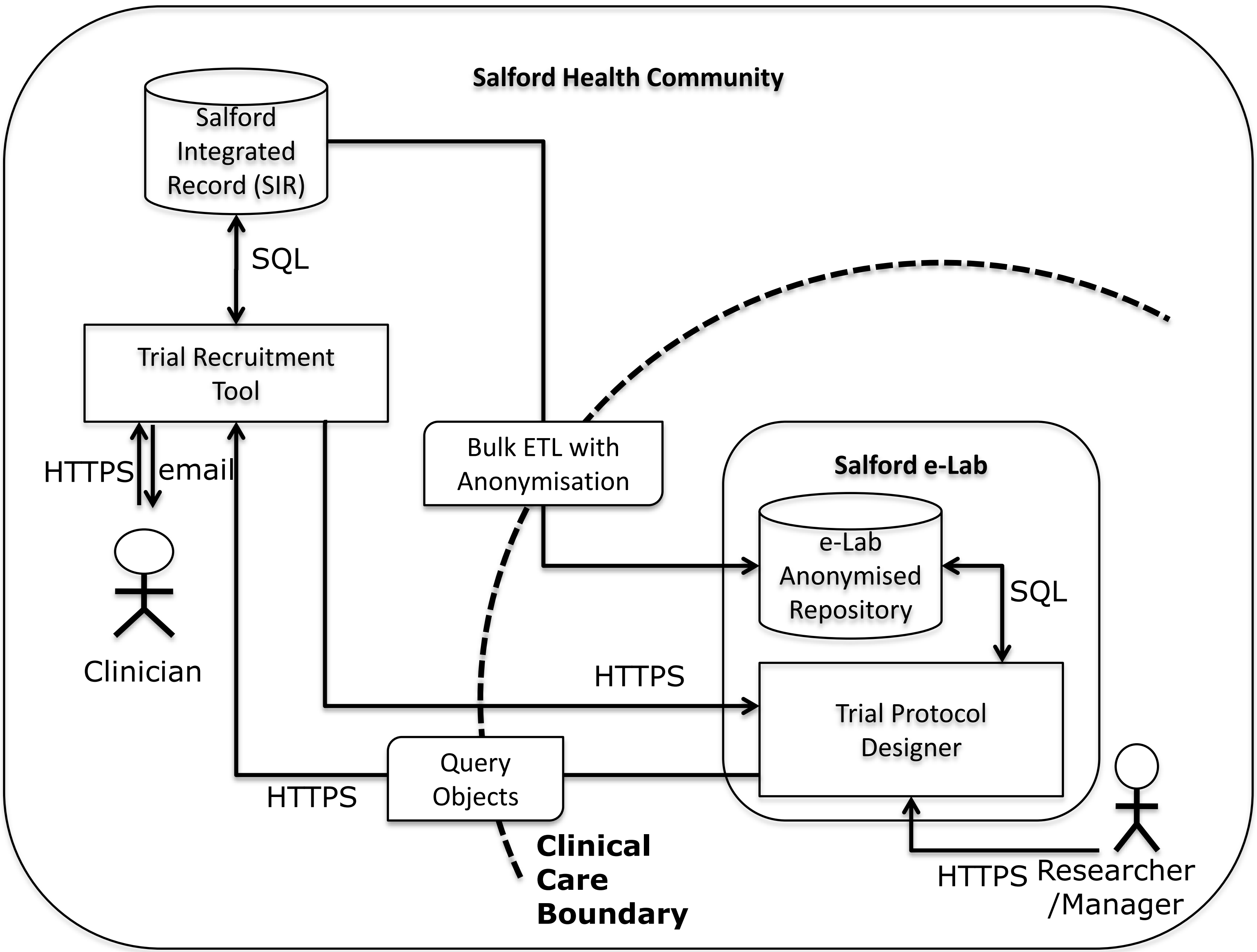
for **I**mproving

Trial **E**fficiency


+/- Tweak
Study Protocol
(sponsor)
Recruitment Estimate
(researcher)
[de-identified records]
Rapid, Consistent
Feasibility & Audit
Refined Search
Study Management
(e-Lab: de-identified)
FIREWALL
Email to attending clinicians
"Click to identify your eligible patients"
[health e-record]
Study Management
(identifiable)

Realistic Recruitment Estimates



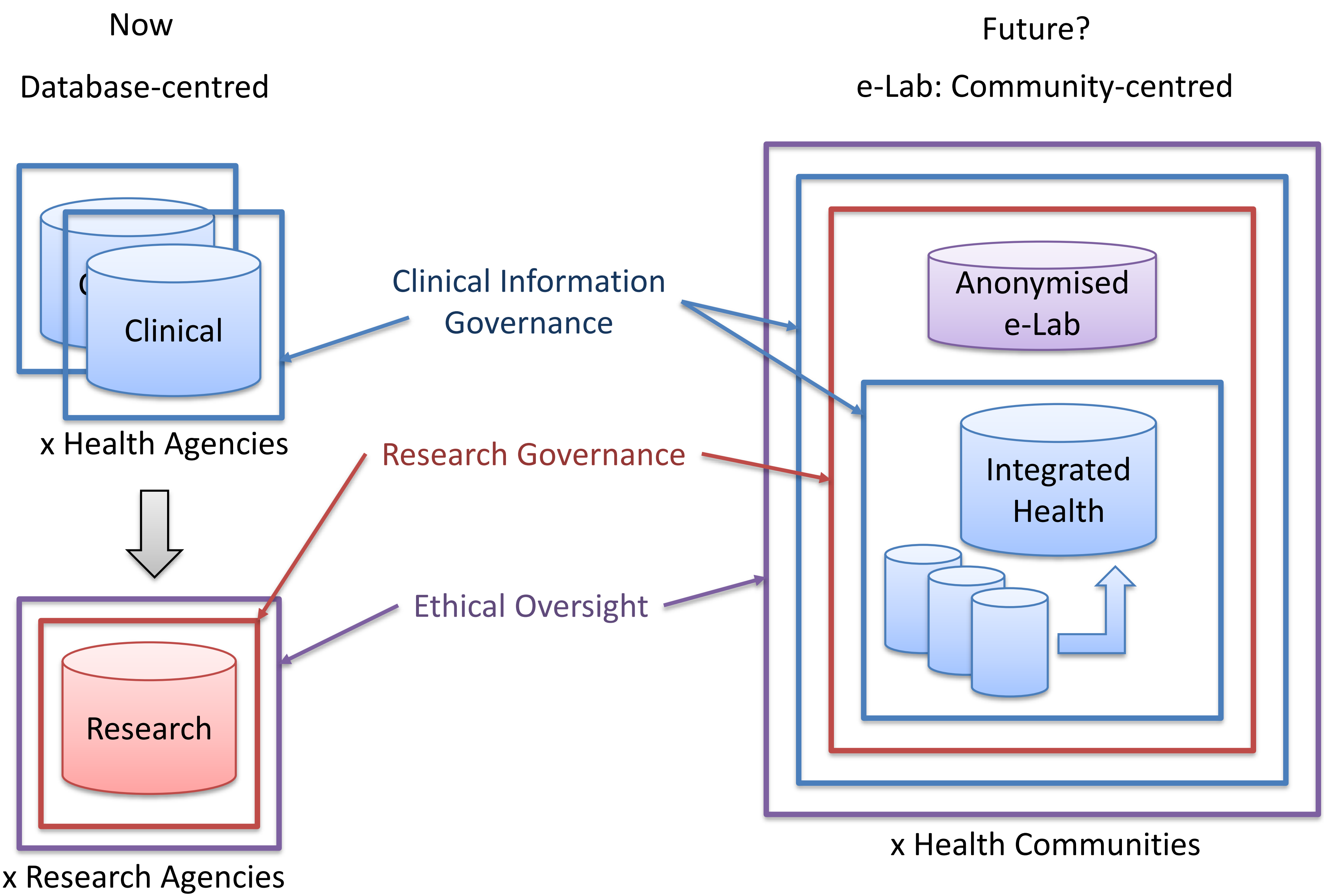


Salford Health Community
Salford Integrated Record (SIR)
SQL
Trial Recruitment Tool
HTTPS
email
Clinician
Bulk ETL with Anonymisation
Salford e-Lab
e-Lab Anonymised Repository
SQL
HTTPS
Trial Protocol Designer
Query Objects
HTTPS
Clinical Care Boundary
HTTPS
Researcher /Manager

Digital Curators Promoting Healthcare Innovation: III

TRUST

Trust & Benefit in Research across Health Records



Public Involvement

- Patient / Citizen Scientist
- Social network ←→ investigator
- Early & mobile signals beyond clinical reach
- Relevant outcomes

DU CONTRAT
SOCIAL;
OU
PRINCIPES
DU DROIT
POLITIQUE.
Par J. J. ROUSSEAU, Citoyen de Genève.

A AMSTERDAM,
Chez MARC-MICHEL REY.
M. DCC. LXII.

Framework for Digital Research in Healthcare: II

SOCIAL CONTRACT

Digital Curation for the Public's Health: What is the Social Contract?

- UK has much already in place in the laws and governance across NHS, higher education and allied public services
- Law, standards, regulations and some infrastructure work nationally
- Trust and capacity-building to provide the best data and analyses works locally ← more attention

NWeH
North West e-Health

Open Unifying e-Lab

$\cup$(open models)

$\cup$

Easy computation

$\cup$

Abstract reasoning and motivation of domain experts

=

More complete insights

More reusable evidence

Better management across diseases

Earlier intervention

Greater citizen involvement