What is an e-Lab? Online social and health data, methods and investigators

Dr. Georgina Moulton

Education and Development Fellow

(georgina.moulton@manchester.ac.uk)

Outline of Talk

- Motivation of e-Labs

- What is an e-Lab?

- Research Objects

- Healthcare e-Labs

- Two examples
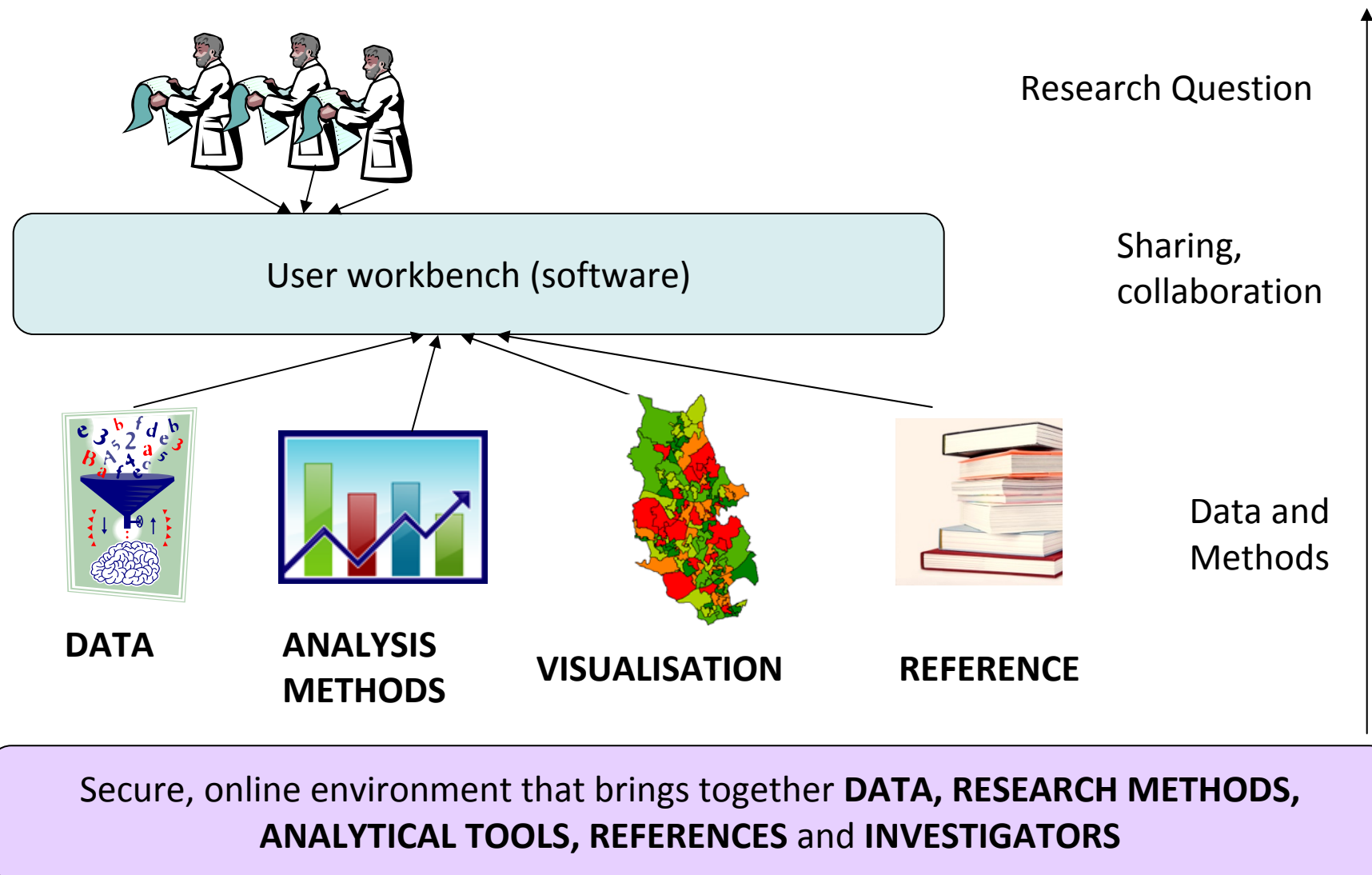 - (1) Obesity e-Lab
  - (2) NWeHealth e-Labs

Motivation for e-Labs

- Silos of research and work
- Divisions occur across
  - multi-disciplinary communities
    - e.g., social scientists and biomedical researchers investigating obesity
  - Within a single discipline community
    - e.g., biomedical researchers investigating nutritional and physical activity components of obesity
- Incomplete research and reinvention of the wheel
- 'Folk Knowledge'
  - Highly contextual
  - Great for rapid knowledge exchange and problem solving
  - Not good for long term repositories or knowledge gain

E-Lab = Knowledge Management System

- ## Knowledge Management
  - Broad group of activities around creation, sharing and adoption of expertise and understanding

- ## e-Lab
  - 'An online, secure environment that brings together data, research methods, analytical tools, references and people at the point of investigation or decision making.'

The *e-Lab* model

Research Question

User workbench (software)

Sharing, collaboration

DATA

ANALYSIS METHODS

VISUALISATION

REFERENCE

Data and Methods

Secure, online environment that brings together **DATA, RESEARCH METHODS, ANALYTICAL TOOLS, REFERENCES** and **INVESTIGATORS**

e-Labs Initiative

- e-Labs serving different communities include:
    - myExperiment
    - SysMO
    - e-Lico
    - Shared Genomics
    - Obesity e-Lab
    - North West e-Health e-Lab initiative

- Parallel development

- Sharing requirements and use cases
    - Building a shared understanding of the architecture required for an e-Lab

- Although implemented in different ways, they share the same aims

What is a Research Object?

- The entities that an e-Lab creates, stores, exchanges, publishes.

- A bundle of digital content
  - Data, methods, parameters, metadata, results

- It's the story of the investigation

- Examples:
  - Laboratory data from instruments, plus blogged lab book entries
  - A collection of all the digital items associated with one experiment
  - A reproducible research article with associated workflows and data

Why use Research Objects?

- Year on year routine analyses
- Rich publication
- Validation
- Extending analyses
- Education
- Collaborating
- Capturing expertise - organisational memory

e-Lab Components

- Technical platform that brings together information
  - Software system/Research Objects
  - Actual information

- Scholarly social networks
  - Attribution, contribution, user engagement, drives content

Health-care e-Labs

Digital Dust (data deposit > use)
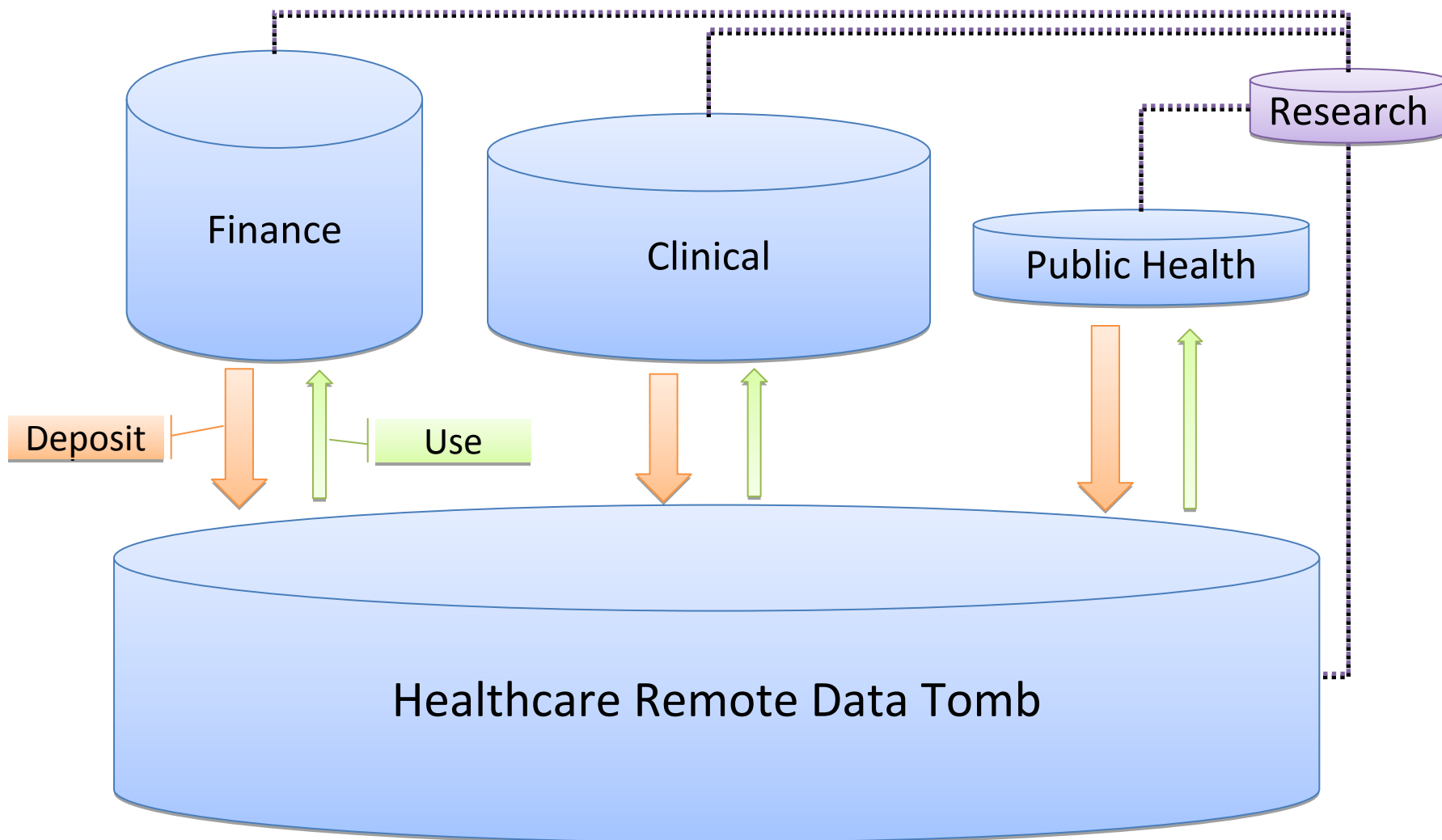
Data supply is not the bottleneck

Methods/Models ↑

Contextual expertise ↔

Data ↑↑↑

Healthcare Knowledge Management

- Healthcare (NHS) native about knowledge management
  - Top down driven networks
  - Mirrors the manner in which policy is implemented
- Successes are often socially driven rather than technically driven
  - CHAIN – developed a community that allowed users to search for areas of expertise
- Knowledge management has increasing importance
 - Economic climate

Current Work Practices

- Primary Care Trusts (PCTs) commission services to meet their local populations (300, 000)

- Analysts who collect and analyse local and national information to support local decision making

- Often belong to informal networks

- Communicate by telephone/email

- Attend training

- Find it difficult to find information

  - Emails!!

- Few links to other working environments or disciplines

Obesity e-Lab: Enabling Obesity Research using the Health Survey for England

Obesity as a Public Health Problem

- More than 1 billion overweight adults, at least 300 million of them obese (WHO 2006)

- Global increases in overweight and obesity are attributable to a number of factors including:
  - Changes in diet – food choices and availability
  - Decreased physical activity – changing nature of work, urban environment, leisure activities…

- Increasing prevalence: by age, sex, ethnicity

- A global phenomenon

- Chronic diseases

- Health costs

- Economic costs

Energy equation balance

Intake

Output

Social determinants

Social determinants

T I M E
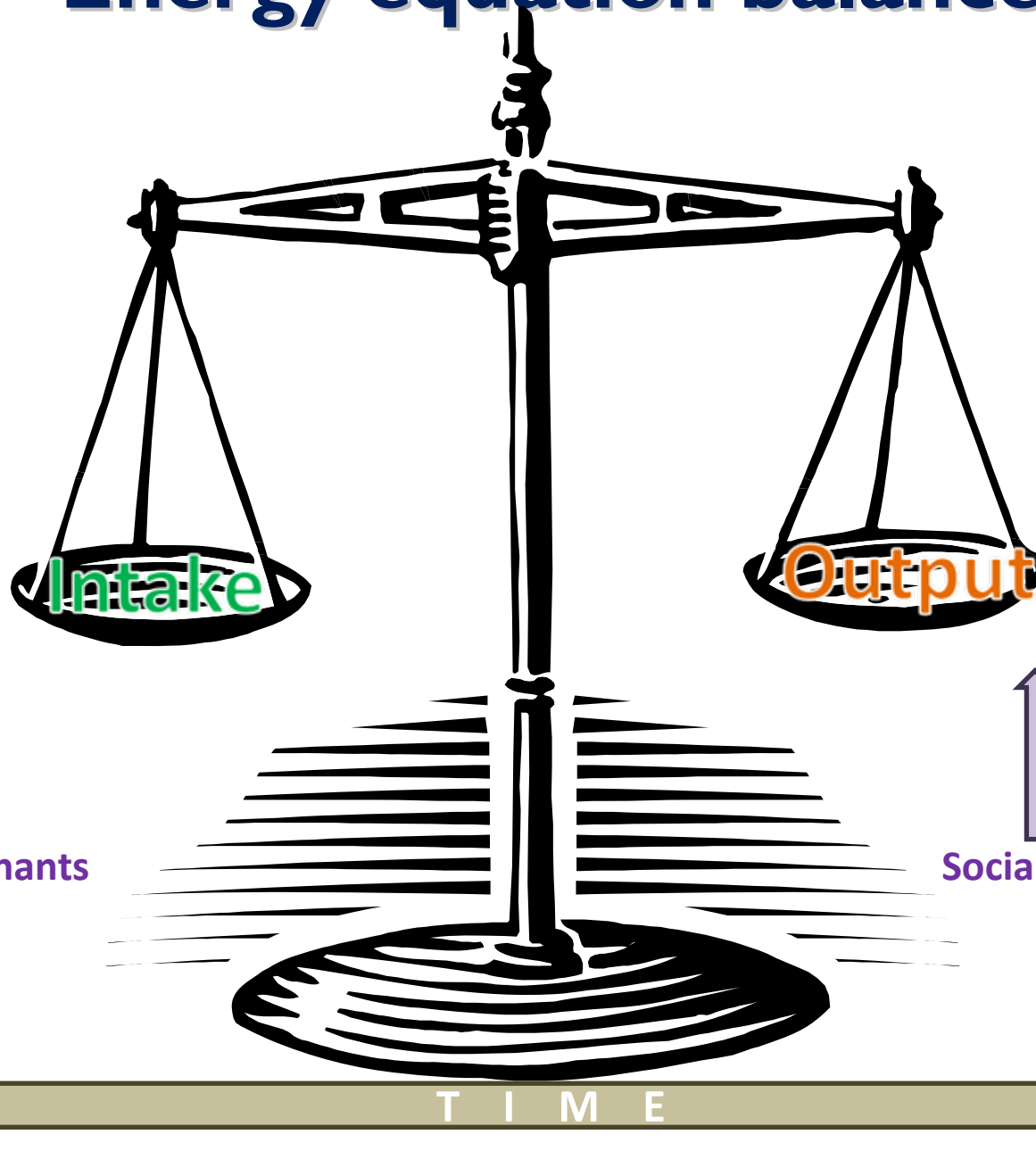
Challenges in Obesity Research

- Multi-disciplinary: requires social, behavioural, bio-medical, public health and environmental perspectives

- Unfamiliarity with other domains of research communities: Lack of understanding of theoretical concepts underpinning the data collected by other disciplines

- Lack of awareness of relevant data sources: Secondary data sources

- Lack of sharing of reproducible packages of research incorporating data and data processes

The Obesity e-Lab project

- ESRC-funded project (Sept 2008 to Aug 2011)

- Investigators/project team from diverse background:
  - Public Health, Informatics, Statistics, Social Science, Epidemiology, Computer Science, Human Computing Interaction, Software Technical Architecture, Software Engineering, Education & Professional Development

- Work alongside other project teams (*e.g.,* the *e*-Science community, text mining community, etc)

- Supported by stakeholders representing various groups (local health authority, leaders of obesity-related initiatives, etc.)

User Communities

- Academic researchers
  - Public Health/Epidemiology/Medicine
  - Social Science

- NHS and Department of Health analysts
  - Local public health decision making
  - Department of Health strategy

Attitudes to sharing

	Data	Scripts
Academic social scientists	Yes	No
Academic epidemiologists/medical researchers	No	Yes
PCT analysts	Yes	Yes

Focus on Health Survey for England

- Country wide survey of nation's health and associated factors such as exercise and diet

- Since 1991

- Commissioned by the NHS, but really under-utilised

- Although you need to register for access to data, it is essentially open for public use

HSE: problems with usage

- Overwhelmed by size of dataset
  - Each year of the survey contains a core set of questions and answers (known as variables)
  - Some years it focuses on a particular topic of interest (e.g. attitudes to exercise, cardio-vascular health)

- It can be difficult to know where to find the variables that would best help address a particular research question.

  - No consistency between years – different variable names, questions change
  - The documentation for each survey is extensive and difficult to search: several hundred pages of documentation per year
  - PDF based
- How to use the data
  - When to apply weightings to survey – NO CLEAR RULES!
- No worked examples

HSE: problems with usage

- Public Health Analysts and Novice Users avoided use to a lesser or certain extent

- Academic users had devised coping strategies

# MethodBox: a platform for exploring survey data

For the individual researcher

- Find and extract survey data of interest
- Search for variables across HSE surveys
- Easier access to documentation
- Make a trail of what you are doing and what you have selected

For the community

- Share expertise and create dialog around the analysis of survey data
- User profiles
- Share scripts, calculations and paper references
- Can decided who sees what – 3 level privacy (you, colleagues, everyone)

http://dl.dropbox.com/u/4958764/methodbox/methodbox_audio.mov

Linking Resources: Example Research Object
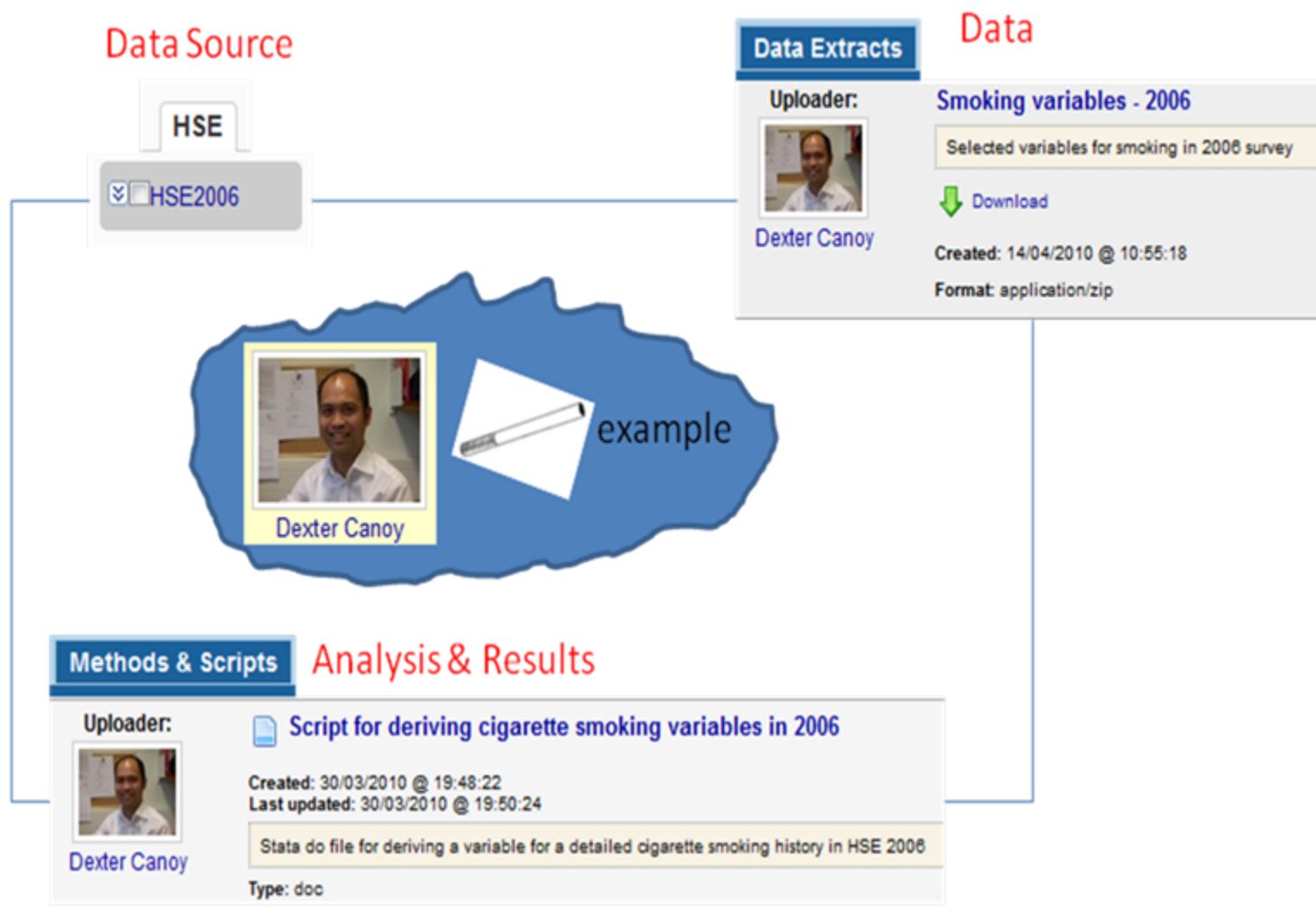
Next steps and long-term vision

- Building User Community
  - NHS Action Learning Sets, PhD supervisors/students, topic-specific research groups

- Provide rich worked examples of analyses from different disciplines

- Continue dialogue and engagement with other users of HSE
  - Is this concept interesting to you?
  - Can we ask about how you use HSE or other survey data?
- Include other survey datasets
  - Welsh Health Survey Dataset
  - Local surveys

NWeHealth e-Lab: transformation of health record data into improved quality and efficient health-care
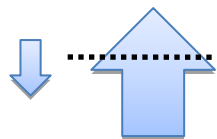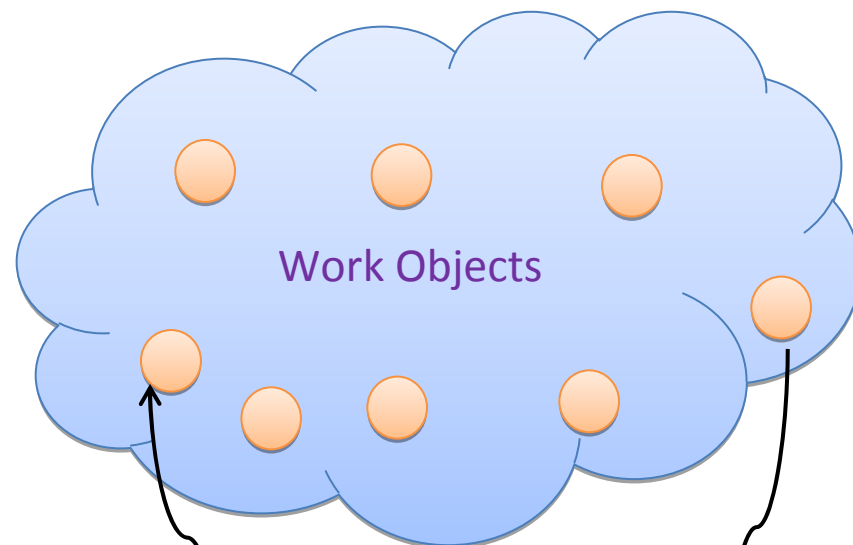
http://www.nweh.org.uk

NWeHealth e-Lab Project

- Collaboratively funded project between Salford NHS, Salford Foundation Trust, University of Manchester and Northwest Development Agency (2008-2011)

- 30 member team with a wide variety of backgrounds
    - Scientific team: epidemiologists, statisticians, mathematicians, knowledge management
    - Clinical team: informaticans, clinicians, public health professionals
    - Software team: computer scientists, software engineers
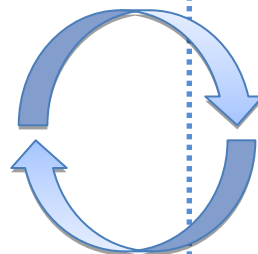   - Operations team: management and operations research, education and development

NWeHealth e-Labs

- Facilitates and stimulates the development of **integrated patient information** data sources

- Allows interrogation, sense-making, access and visualisation of data

- Ensures that all data access and handling is **secure and compliant with best practice**

- User population is larger and more diverse
  - Health economists
  - Performance managers
  - Commissioners
  - Public Health Analysts/Managers
  - Clinicians

NWeHealth e-Labs

Smart Summary Views of Care

Commissioning Intelligence

Research Assets

Clinical Audit Intelligence

North West e-Health e-Lab: Analytical Layer

Connecting for Health LSPs & ESPs: Core Services

Turning INFORMATION into RESEARCH/BUSINESS/PUBLIC HEALTH/CLINICAL INTELLIGENCE

Health care Intelligence and Decision Support Making Questions

- **Deaths**
  - Life expectancy for Trust and wards
  - Overlay the deprivation scores
  - Overlay areas of high clinical deprivation e.g. diabetes, CVD

- **Births**
  - Trend in birth rates by Ward
  - Proportion of new mothers by age / ward who have an admission during pregnancy
  - Which mothers are at risk of gestational diabetes?
  - Which mothers have LTC

- **Lifestyle factors**
  - Hospital admissions for smokers / non-smokers
  - Height and weight (Body Mass Index) by age ward

- **A&E attendances**
  - How many people with LTC attend A&E? What is cost of this?
  - How many older people use A&E?

- **Registers**
  - Are people taking statins on a QoF register?
  - What are the outcomes of early detection/intervention in comparison to the rest of the popn?
  - What are the costs for a diabetic (primary & secondary care)?
  - Graph post operative wound infection rates by speciality
  - Total cost of treatment for patients with LTC

- **Performance**
  - Healthy lifestyle advice given to patients with hypertension diagnosed after 01/04/09
  - CHD patients with a BP reading of 150/90 or less
  - Teenage Pregnancy rates (age 15-17)
  - Vaccination completion rates by recommended age

Examples of Research Objects

- Template equity audit
  - Workflows, instructions, default data and tutorial material

- Research article describing workflows and data required to reproduce results

- Reproducible annual PH report
  - report and supporting files such as slides, spreadsheets and graphics
  - commentary around the report such as copies of emails and press releases/articles
  - pointers to data sources and extracts used in analyses
  - data extraction, cleaning, derivation and statistical analysis workflows behind the analyses
 - bibliography; project management resources.

Information Governance

- Patient data
  - highly sensitive subject
- Owner of data always with patient
- Custodians of data are GPs, PCTs, Caldicott Guardians…. persons of responsibility who 'host' and use data
- Rules of ownership stay the same in an e-Lab
- Information Governance Framework that manages access to data
  - Security, confidentiality, risk management
- Information Governance/Ethics Board
  - Delegated responsibility from National Research and Ethics Committee

Example NWeHealth e-Lab
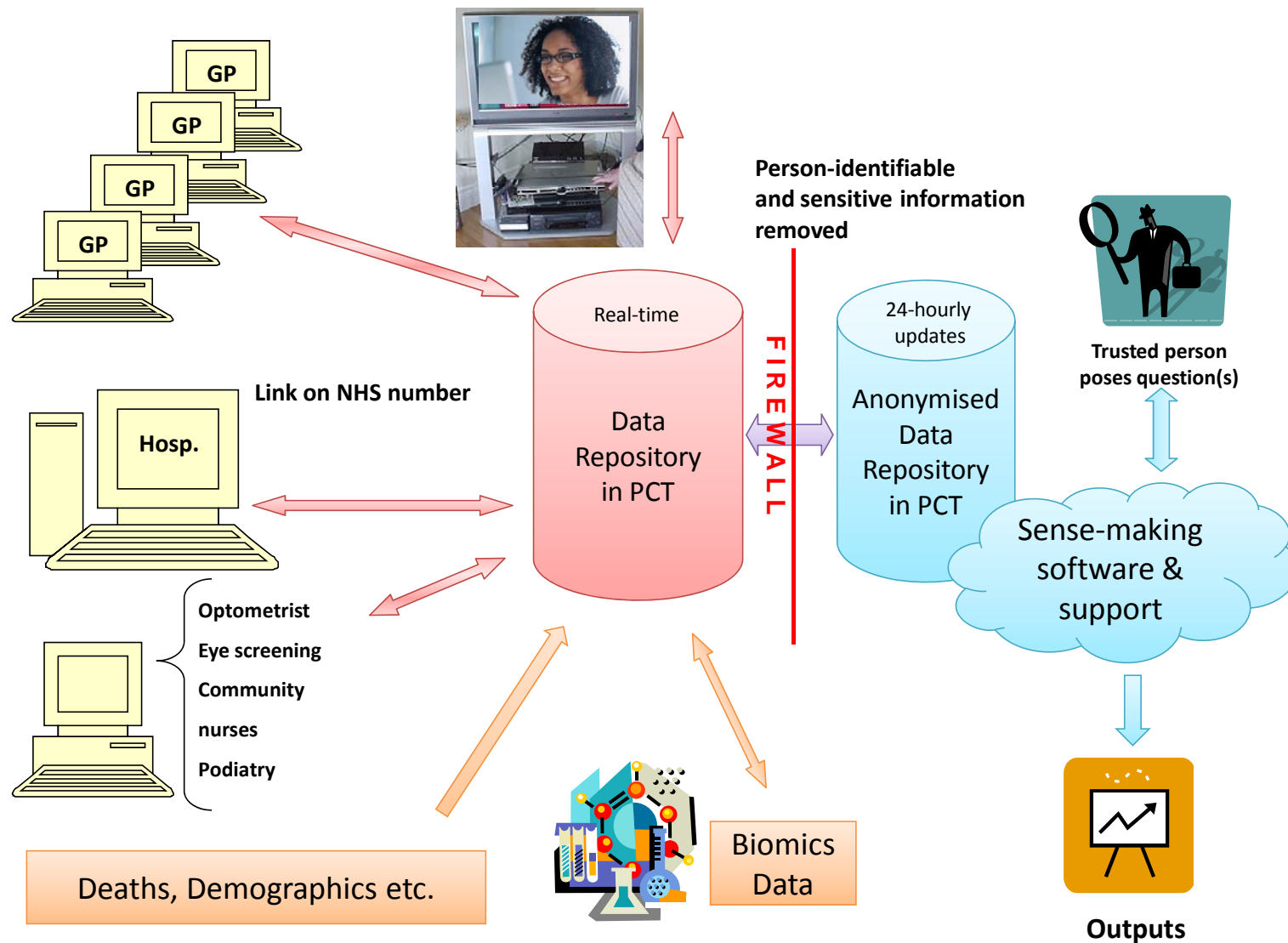
Federation of e-Labs
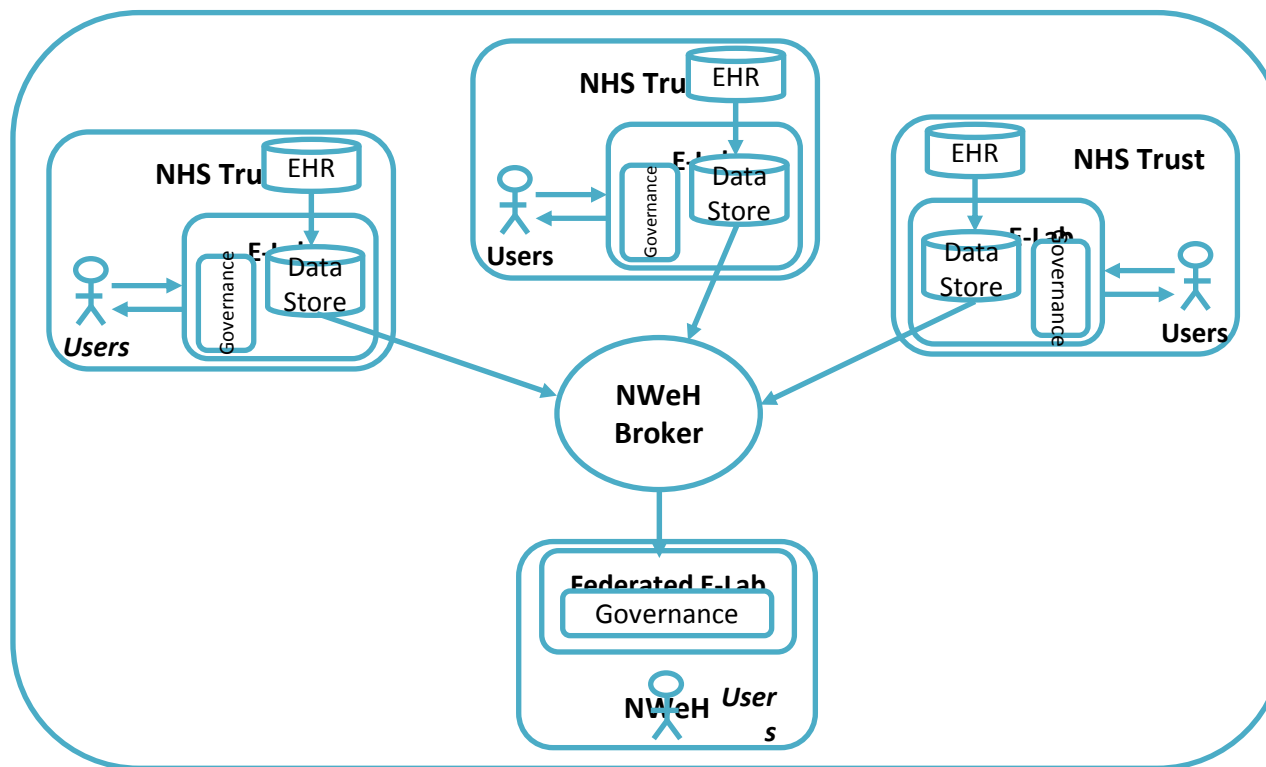
The e-Lab platform can be **replicated across all NHS Data Partners** and **access to data aggregated**

Advantageous: **complete picture of a patients journey**

e.g., individual or population level – post-operative care/community-care

Prototype NHS e-Lab

Life Expectancy In Salford

settings

Data

Project Details
Data
Documents
Data Exploration
Notes
Snapshots
People

Help

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	DiabetesPreva	CvdPrevalenc	DiabetesPrevi	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

settings

Data

Project Details

Data

Documents

Data Exploration

Notes

Snapshots

People

Help

Life Expectancy

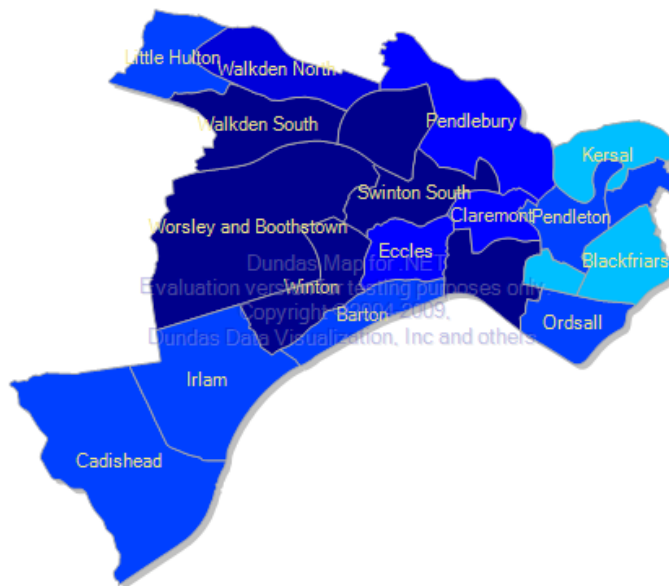(Show Details) (Add a Note...) (Add a File...)

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

map   settings

Take Snapshot

Little Hulton
Walkden North
Walkden South
Pendlebury
Kersal
Swinton South
Claremont Pendleton
Worsley and Boothstown
Eccles
Blackfriars
Winton
Barton
Ordsall
Irlam
Cadishead

Dundas Map for .NET
Evaluation version, testing purposes only
Copyright 2004-2009,
Dundas Data Visualization, Inc and others

Legend Title
- 69 - 70
- 70 - 70
- 70 - 73
- 73 - 74
- 74 - 75
- 75 - 79

Future Work

- To include knowledge management practices into software development – Continued Professional Development
  - Use research objects as teaching examples in Action Learning Groups

- Accreditation through CPD points that map onto key competencies

- Provenance trails
  - Important in relation to Information Governance

Summarise Health-care e-Labs

Obesity e-Lab

- Defined problem and area
- User Population from different disciplines and working environments
- Data is structured and systematically collected
- Data is open access after registration

NWeHealth e-Lab

- Diverse range of uses – clinical audit, business intelligence, clinical pathways
- User populations have greater discipline diversity
- Data is 'messy'
  - Not recorded in a systematic way
  - Often requires cleaning and manipulation
- Data requires protection

e-Lab Conclusions

- Implemented for a range of **different communities and disciplines**

- Implemented for **different purposes**

- The way it is implemented will be different depending on user community requirements

- e-Labs share **common components/aims**:
  - Technical platform
  - Social networking

- Will enhance existing or develop new research/work by **improving knowledge transfer**

- Shift in **technical and cultural change** (e.g., working practices, sharing of information
  - **e-Labs** and **Research Objects** are **not familiar concepts**

Acknowledgements and References

- Obesity e-Lab team: **Iain Buchan**, Carole Goble, Angela Dale, Dexter Canoy, **Sarah Thew**, Shoaib Sufi, Ian Dunlop, Urara Hiroeh

- NWeHealth: Iain Buchan, Stephanie Austin, John New, Burrinder Grewal, Patricia Baker, Mike Shinks, **Bruce Maghill**, Sara Muller, Eleanor Bannister, Milena Falcaro, Gary Leeming, Robert Harper, Lucy Bridges, Paul Jarvis, James Cunningham, Peter Lackey, Owen Williams, Shaun Stanley, Gillian Gaskell, Claire Panes, Amira Thorn

- *Thew, S., Jarvis, P., Ainsworth, J. and Buchan, I* Obesity Atlas and Methodbox: Towards an Open Framework for Sharing Public Health Intelligence Workflows. *MedInfo Conference 2010*

- *Ainsworth J, Buchan I.* e-Labs and Work Objects: Towards Digital Health Economies. *In Communications Infrastructure. Systems and Applications in Europe. Eds. Mehmood R, Cerqueira E, Piesiewicz R, Chlamtazc I. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer, Berlin Heidelberg 2009; 16: 206-216. DOI: 10.1007/978-3-642-11284-3.*

Thank you! Any Questions?