University of Leicester

JISC

NHS National Institute for Health Research

BRISSKit update to NHS-HE Forum, May 2013:

Biomedical Research Infrastructure Software Service Kit

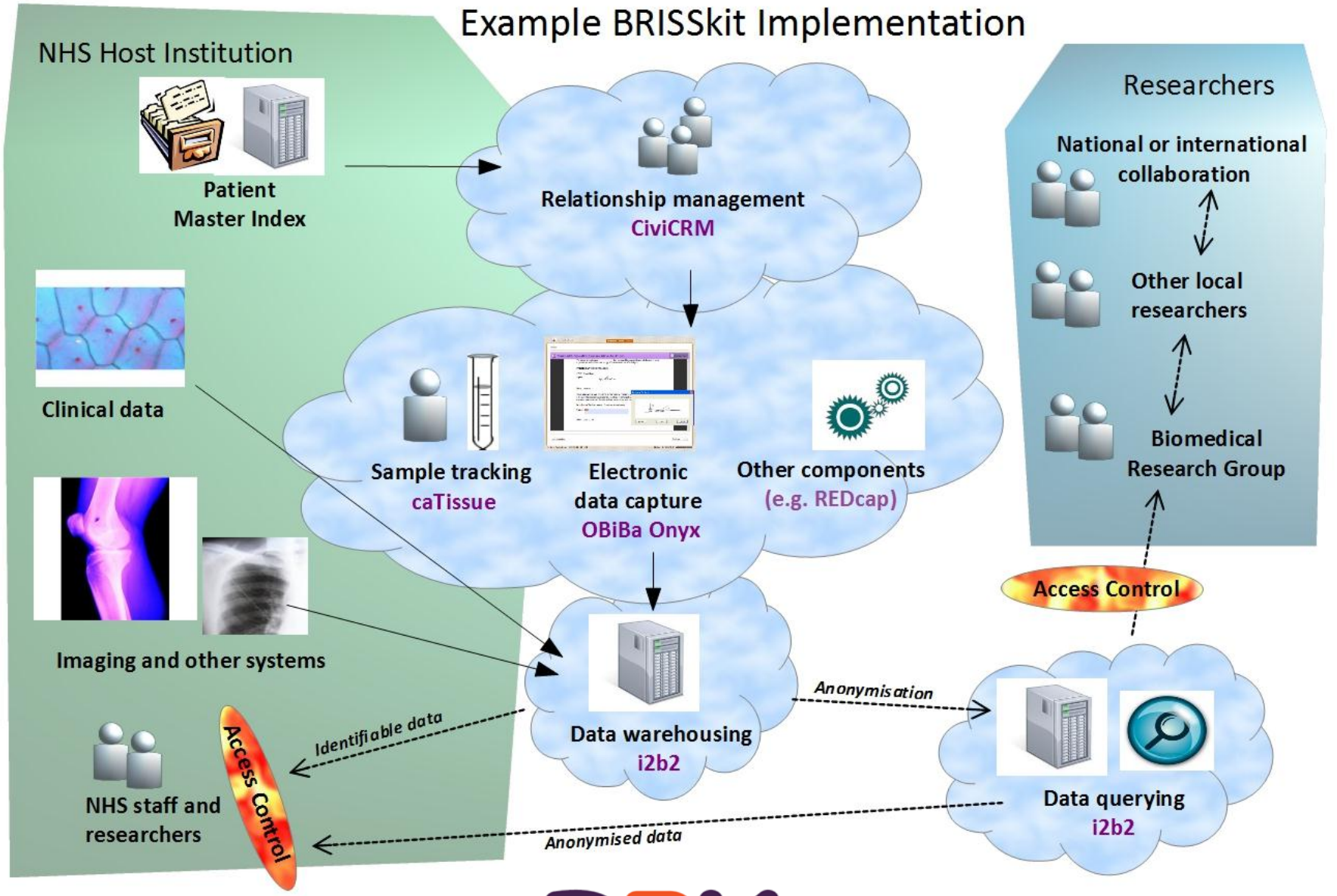
A vision for cloud-based open source research applications

Dr Jonathan Tedds (Project Lead, Senior Research Fellow, D2K Health Sciences)
Dr Andrew Burnham (Information Governance Lead, IT Services)

http://www.brisskit.le.ac.uk

D2K data to knowledge

Example BRISSkit Implementation



D2K
data to knowledge

BRISSKit USPs

- Integrated support for core research processes
- Well-established mature open source applications as protoyped in Cardiovascular: fully UK customised
- A platform for seamless management and integration between applications
- An API allows integration with existing clinical systems
- Easy set up, use and administration through browser (including on mobile devices)
- Capability of being hosted in any compliant cloud provider including UHL (NHS information governance)

BRISSKit Community & Hack Event, Oct 2012

BRISSKit Community & Hack Event

- http://www.brisskit.le.ac.uk/node/35
- created ideas pre and post event via healthresearchhack google group
- 6 hack solutions in 2 days using BRISSKit stack, e.g.
  - i2b2 integration using demo data from HES and cancer research clinical trials data (UCL, Birmingham, Goettingen, Leicester)
  - smartphone app to scan v.tiny barcodes from the end of sample vials and import info into caTissue
  - integrate CiviCRM study management and REDCap questionnaire tool (UHL Respiratory BRU)
  - create a simple CiviCRM study creator as a Drupal plugin

www.brisskit.le.ac.uk

SEARCH

Email: *brisskit@le.ac.uk*

BRISSKIT

Biomedical Research Infrastructure Software Service Kit

HOME | ABOUT US | PRODUCTS & SERVICES | CASE STUDIES | EVENTS | BLOGS | WIKI | CONTACT

BRISSkit - Biomedical Research Infrastructure Software Service kit

Overview

BRISSkit will design a national shared service brokered by JANET to host, implement and deploy biomedical research database applications that support the management and integration of tissue samples with clinical data and electronic patient records. We are uniquely positioned to tackle this through our experience in developing the pioneering open source IT infrastructure for the Biomedical Research Informatics Centre for

JONATHAN TEDDS

- My account
- Create content
- Log out

RECENT BLOG POSTS

- CiviCRM
- Advances in Medical Sciences

University of Leicester

JISC

D2K data to knowledge

University Hospitals of Leicester NHS NHS Trust

NEW: BRISSKit and Bio Banking

- Deploy solutions in international bio banking initiatives
- Investment through Prof Paul Burton (Health Sciences at Leicester/Bristol) & international collaborations
- Building on strong informatics expertise at University of Leicester in partnership with the University Hospitals Leicester Trust
  - Cardiovascular, Respiratory & Lifestyle BRUs
  - Genomics etc

Large data sets, why bother?

THEORY AND METHODS

Size matters: just how big is BIG?

Quantifying realistic sample size requirements for human genome epidemiology

Paul R Burton,[1,2,3*,†] Anna L Hansell,[4,†] Isabel Fortier,[3,5] Teri A Manolio,[6] Muin J Khoury,[3,7] Julian Little[3,8] and Paul Elliott[4]

International Journal of Epidemiology 2009;38:263–273
doi:10.1093/ije/dyn147

- Sample size
- Depth of phenotyping
- Quality of measurement

All critical

BioSHaRE.eu

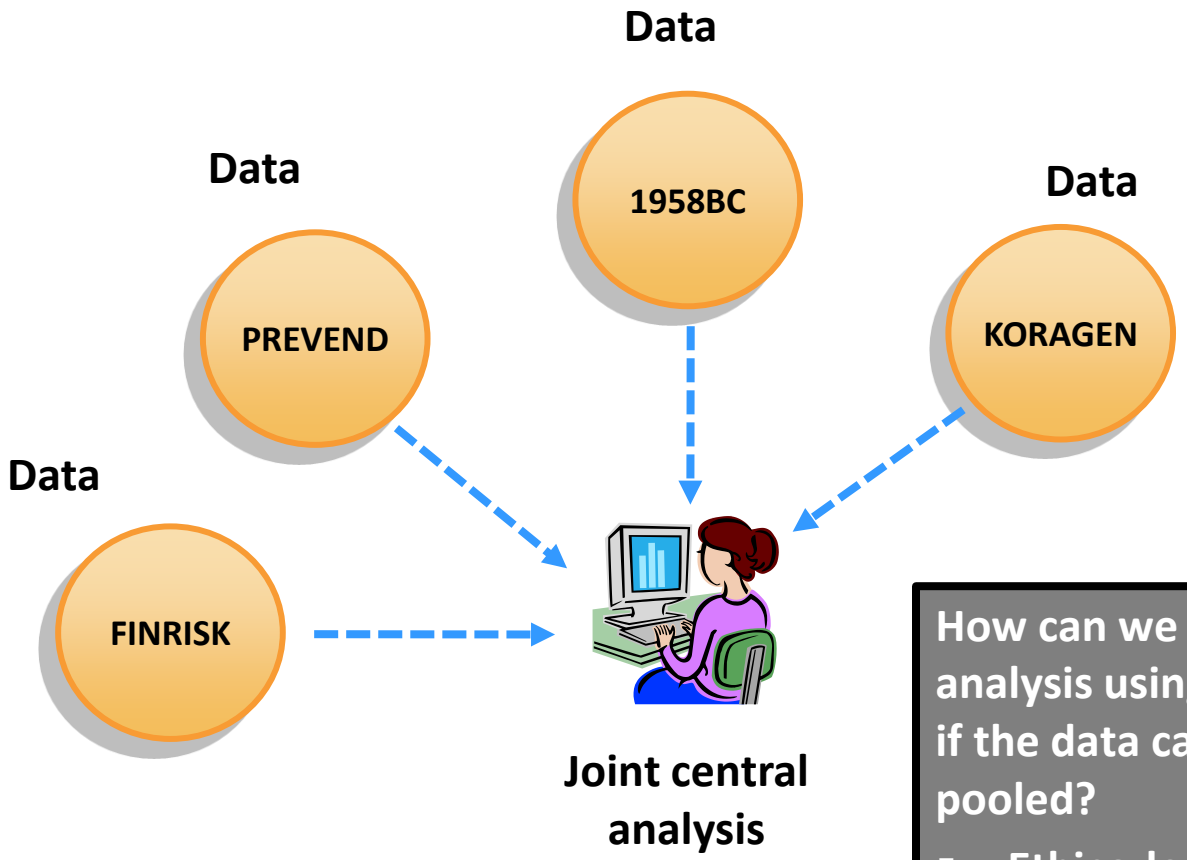
D2K data to knowledge

How big is BIG?

- The direct effect of a gene
  - 2,000 cases minimum, 10,000 cases better
- Environmental and life-style factors
  - Highly context specific: from hundreds to tens of thousands of cases
- Gene-lifestyle and gene-gene “interactions”
  - Absolute minimum 10,000, usually need at least 20,000, a comprehensive platform needs at least 50,000
  - Scientifically fundamental

COOPERATION
BioSHaRE.eu
D2K data to knowledge

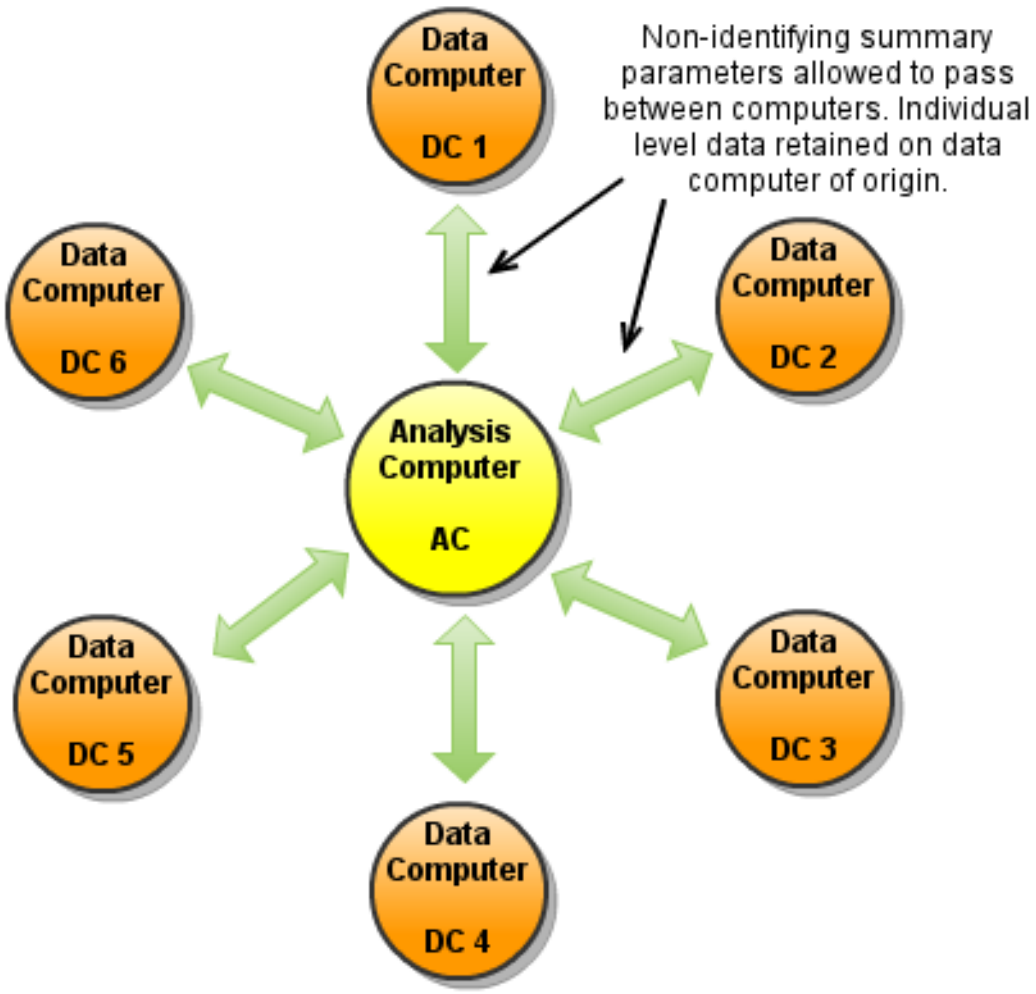
Horizontally partitioned data



How can we undertake a full joint analysis using multiple data sources if the data cannot physically be pooled?

- Ethico-legal constraints
- Physical size of the data objects
- Intellectual property issues

DataSHIELD: a novel solution



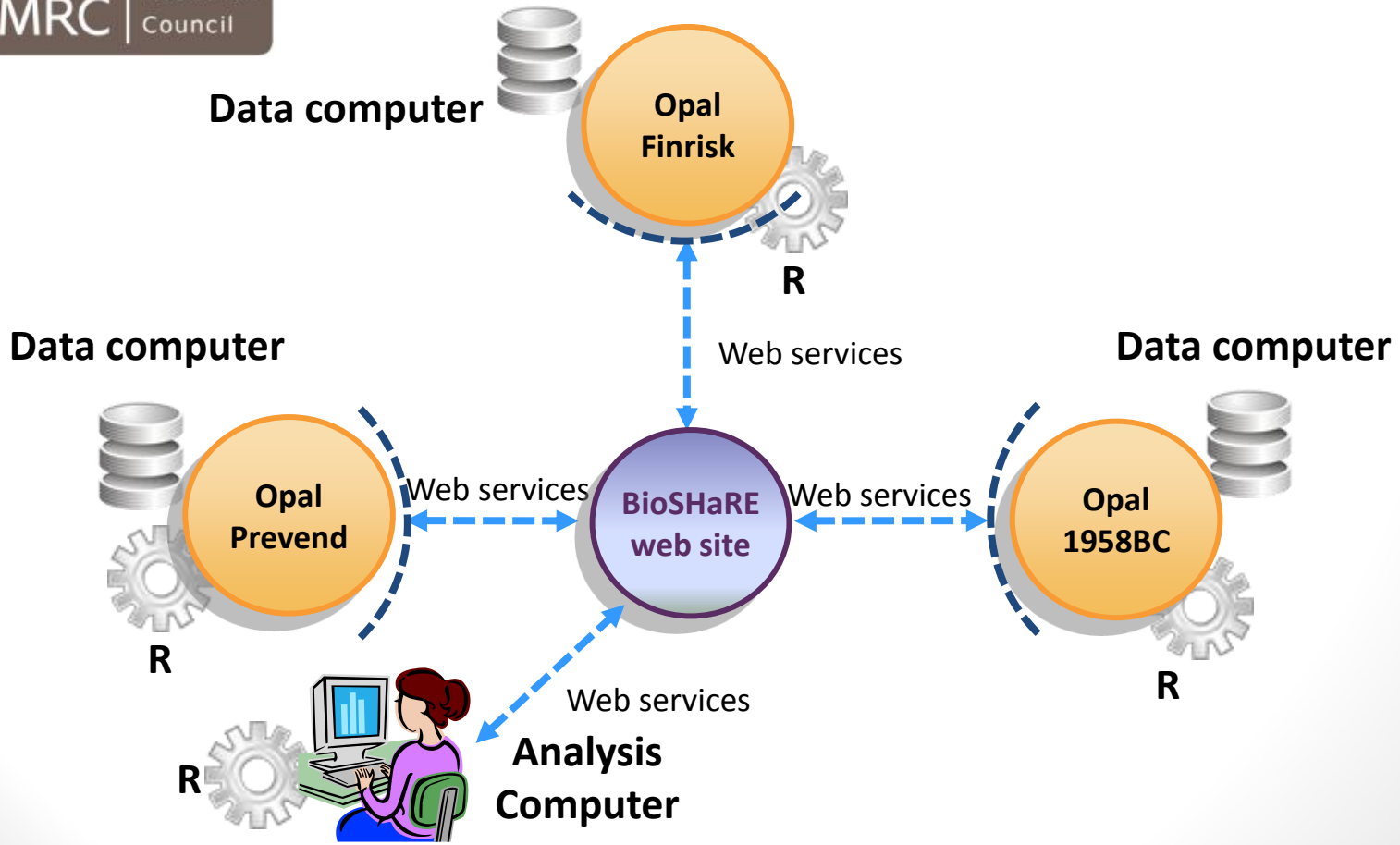
Take analysis to data not data to analysis

One step analyses: simple

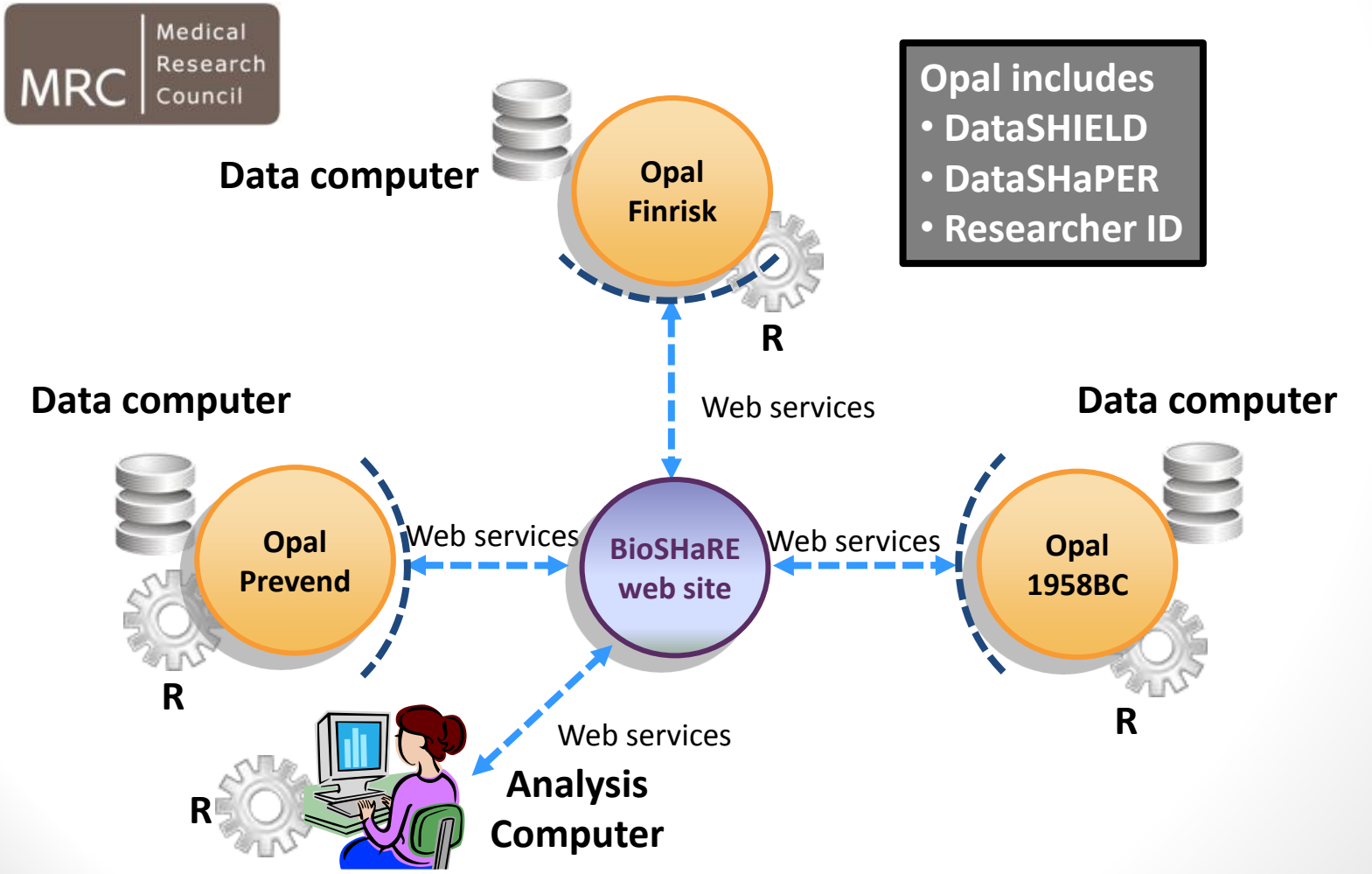
Iterative analyses: parallel processes linked together by entirely non-identifying summary statistics

Typically produces mathematically identical results to fitting a single model to all the data held in one pooled data set

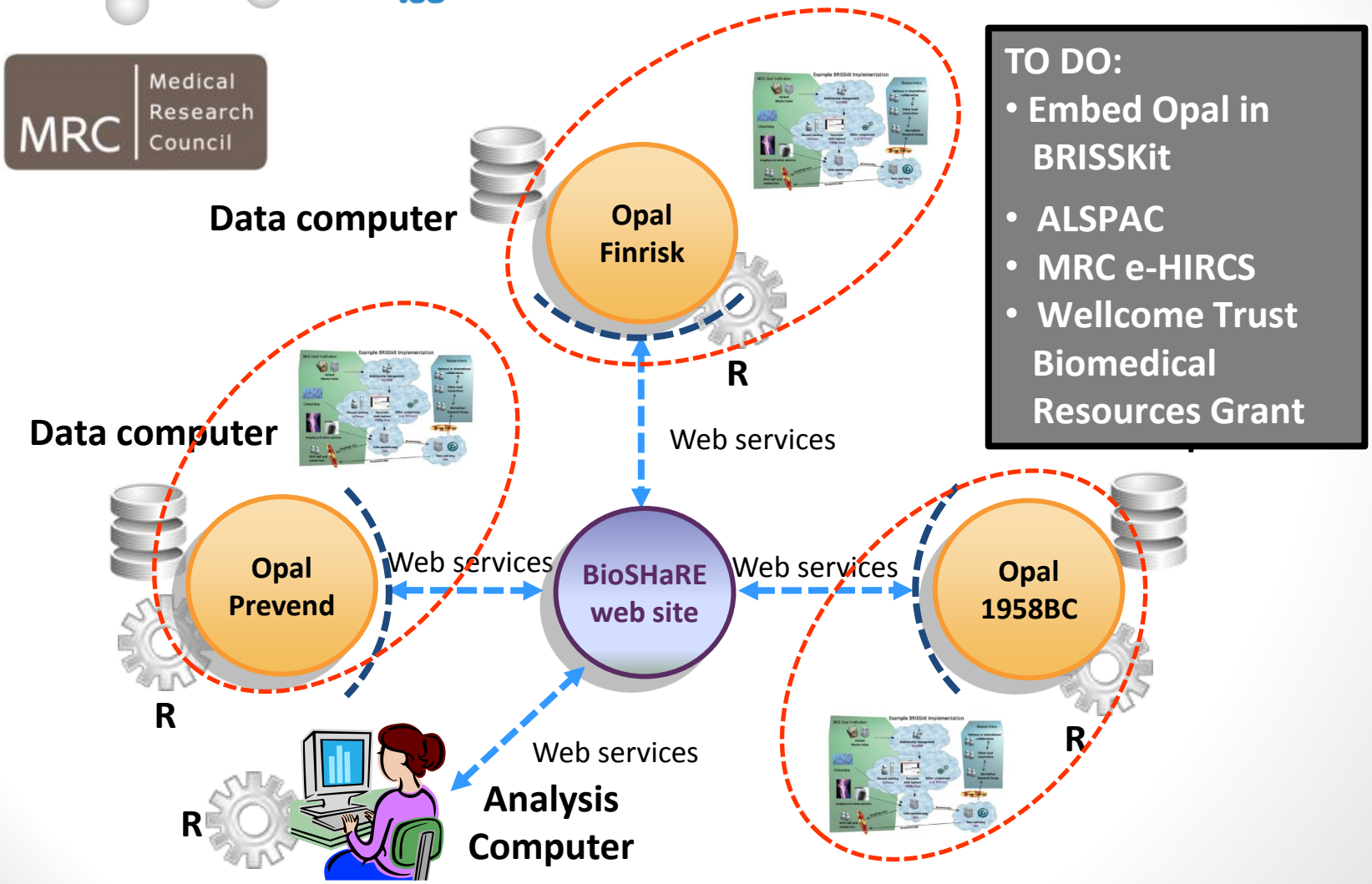
Horizontal DataSHIELD



Horizontal DataSHIELD

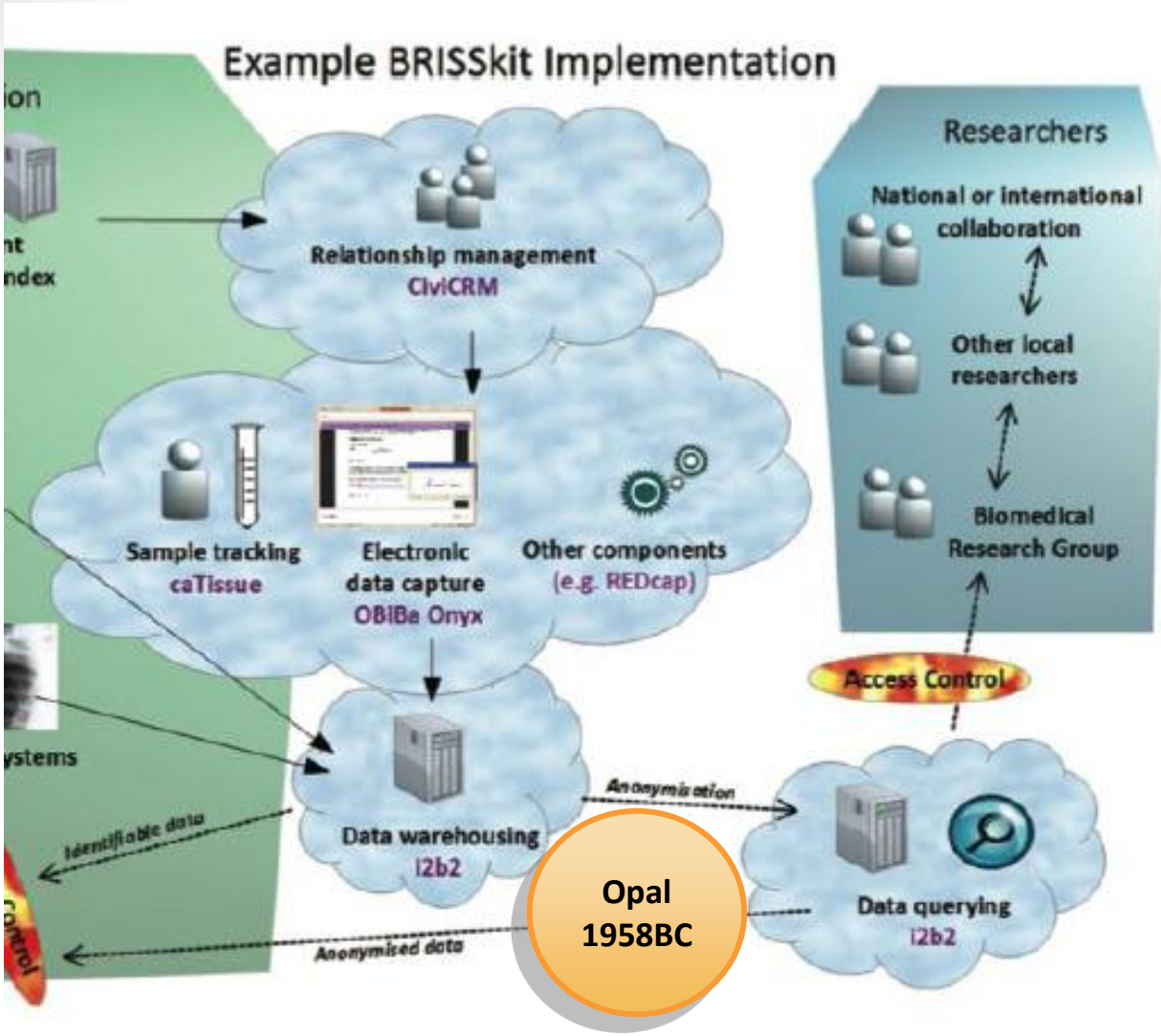


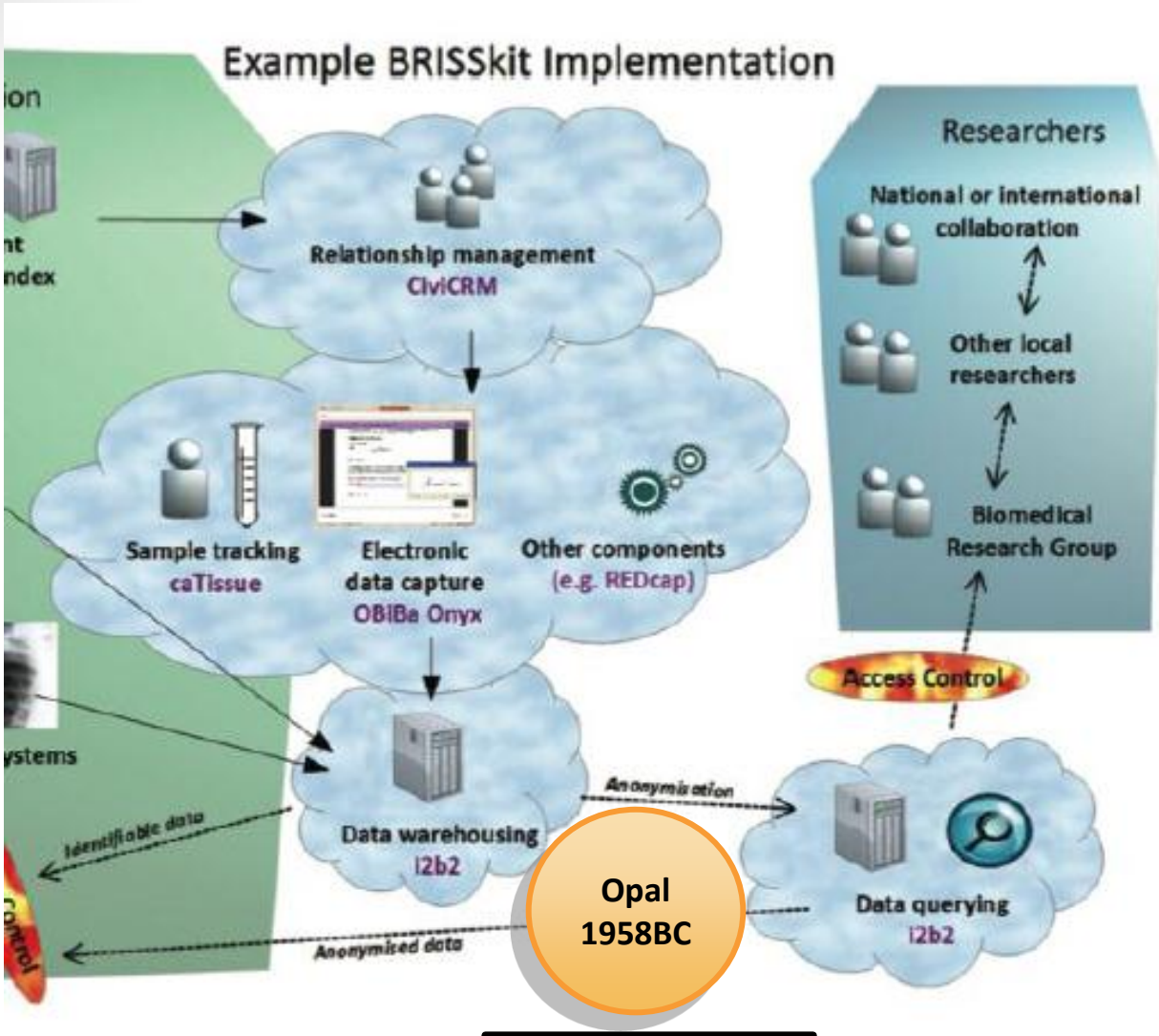
Horizontal DataSHIELD



TO DO:

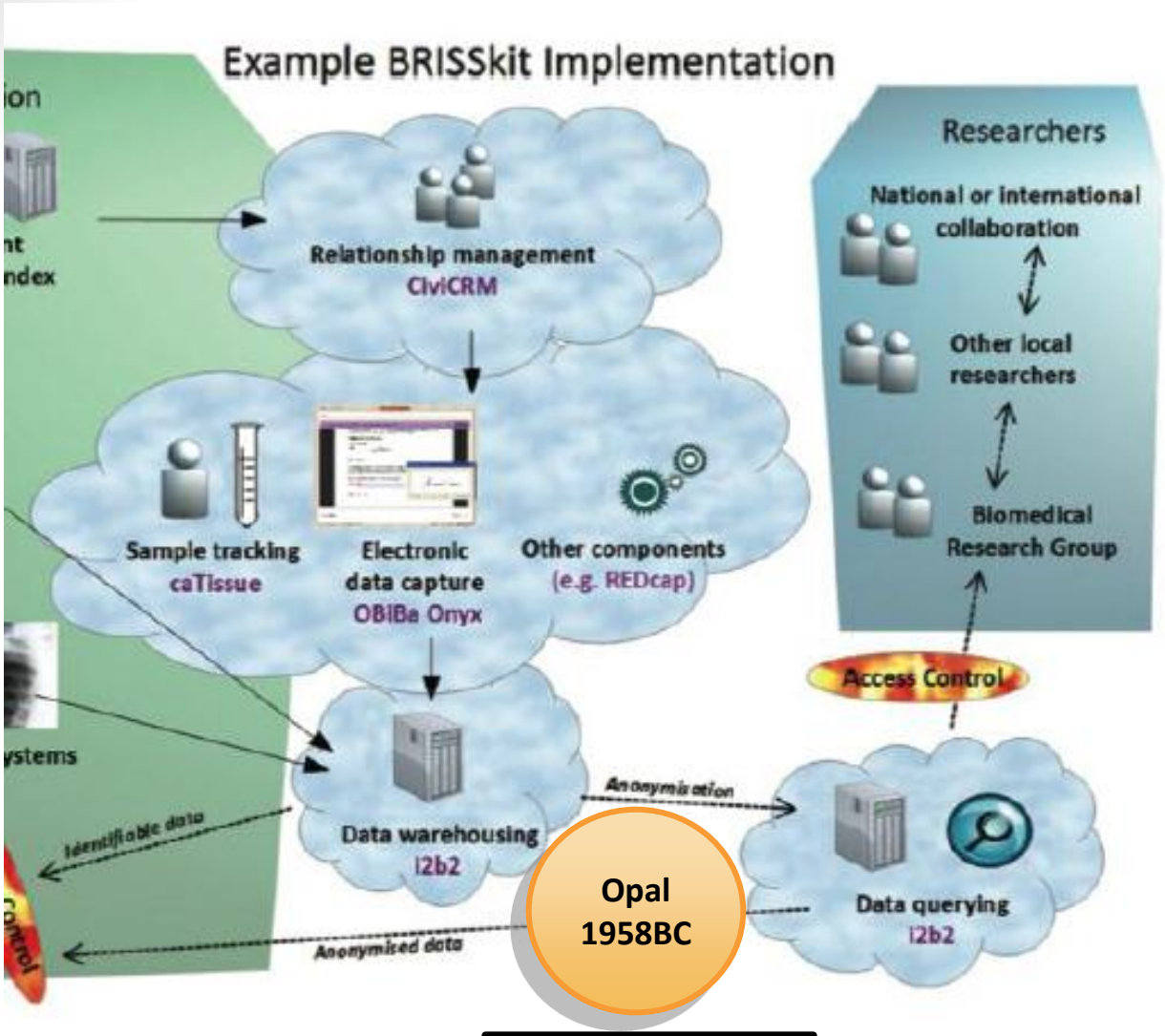
- Embed Opal in BRISSKit
- ALSPAC
- MRC e-HIRCS
- Wellcome Trust Biomedical Resources Grant


Example BRISSkit Implementation
Researchers
National or international collaboration
Other local researchers
Biomedical Research Group
Relationship management
CiviCRM
Sample tracking
caTissue
Electronic data capture
OBiBa Onyx
Other components
(e.g. REDcap)
Access Control
Data warehousing
i2b2
Anonymisation
Identifiable data
Anonymised data
Data querying
i2b2
Opal
1958BC



BRISSKit gains

- DataSHIELD
- DataSHaPER
- Researcher ID

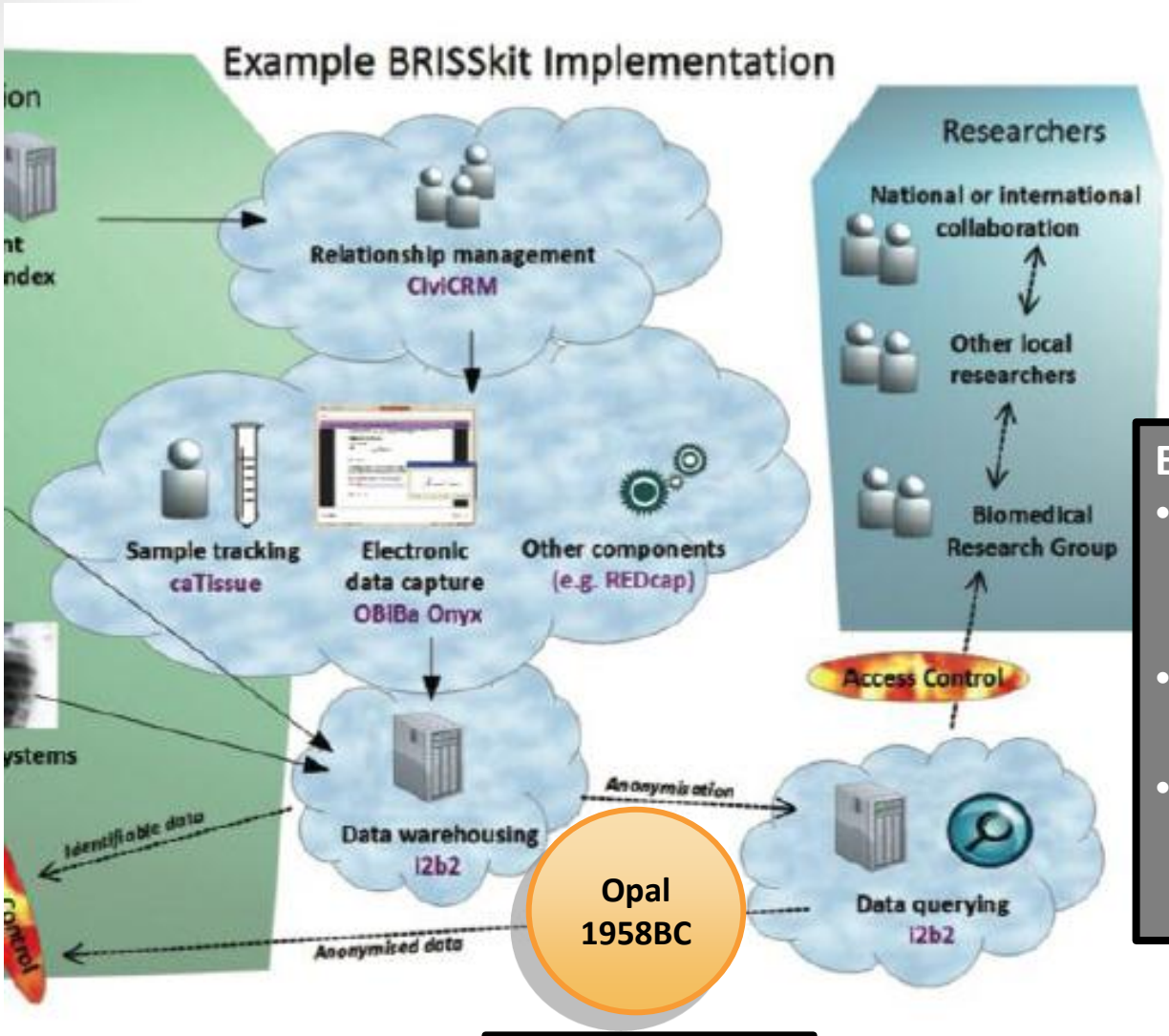


Opal gains

- Direct interface with more tools
- I2B2 functionality
- Potential for enhanced user interface

BRISSKit gains

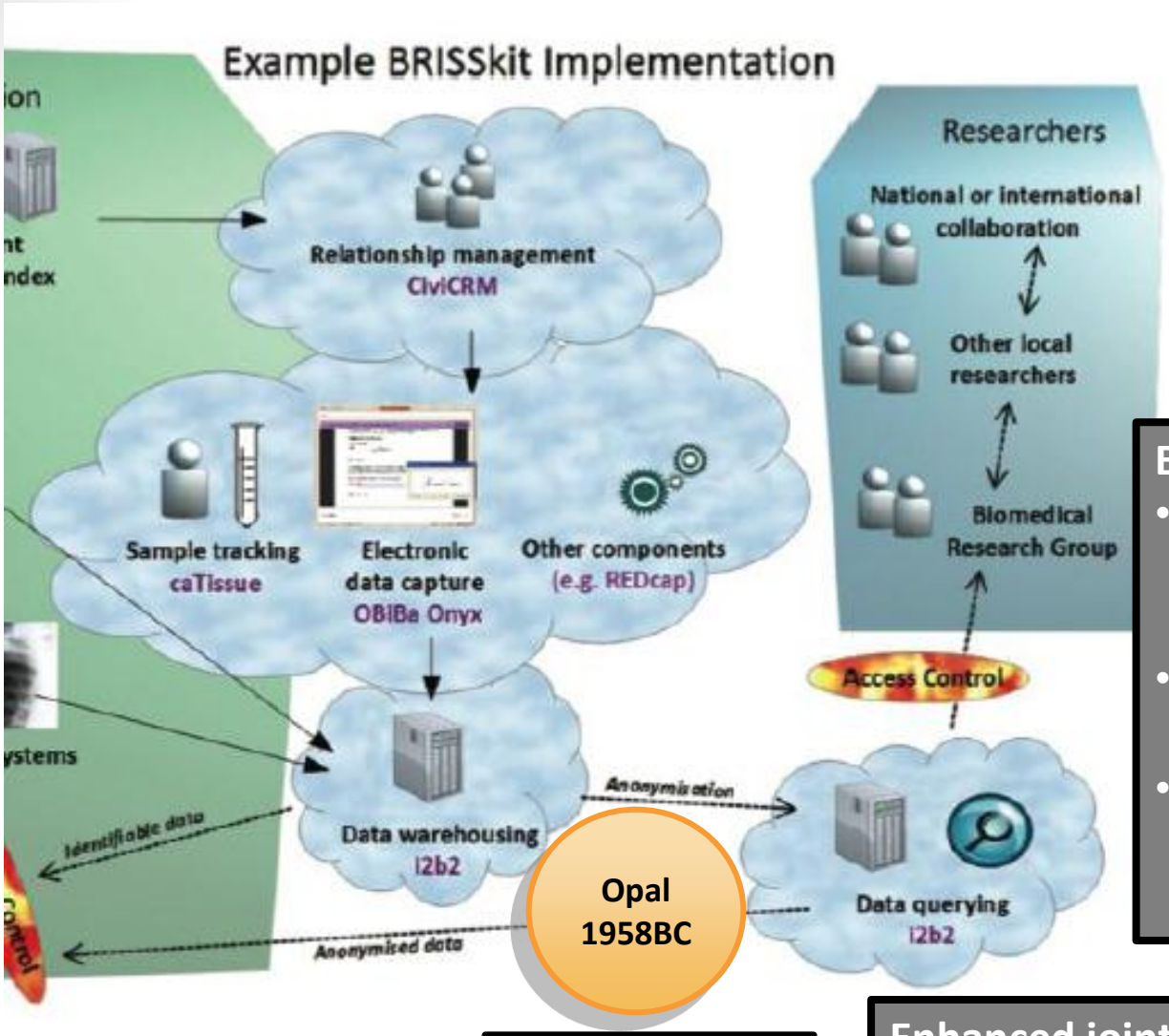
- DataSHIELD
- DataSHaPER
- Researcher ID



Opal gains
- Direct interface with more tools
- I2B2 functionality
- Potential for enhanced user interface

Everybody gains
- Enhanced combined functionality
  - better science
- Bigger user group
  - greater portability
- Greater potential to become a sustainable standard

BRISSKit gains
- DataSHIELD
- DataSHaPER
- Researcher ID



Opal gains
- Direct interface with more tools
- I2B2 functionality
- Potential for enhanced user interface

Everybody gains
- Enhanced combined functionality
  - better science
- Bigger user group
  - greater portability
- Greater potential to become a sustainable standard

BRISSKit gains
- DataSHIELD
- DataSHaPER
- Researcher ID

Enhanced joint analysis with
- Ethico-legal constraints
  *e.g.*US/Europe biobanks
- Intellectual property issues
  e.g. H3AFRICA

The bottom line

- Effective data access is crucial
- Effective joint analysis is essential too (integration)
- Fundamental challenges
  - Scientific harmonization
  - Restriction on access to individual level data
  - Streamlined access to multiple data sets
- Central to the integrative aims of P^3G, PHOEBE, BioSHaRE-eu *etc*
- Also fundamental to the aims of potential BRISSKit users

BRISSKit as a consortium & service

Modular approaches and scalable tools with open source licenses make good investments

- *Individual researchers and associates*
  - enterprise-level tools without the IT overheads
- *Research themes and departments*
  - stand-alone instances of required tools to accelerate research
- *Research units and centres*
  - integrated toolkit with clinical data loading services, or 'jigsaw pieces' to complement existing provision
- JISC internal business case *under construction*…