

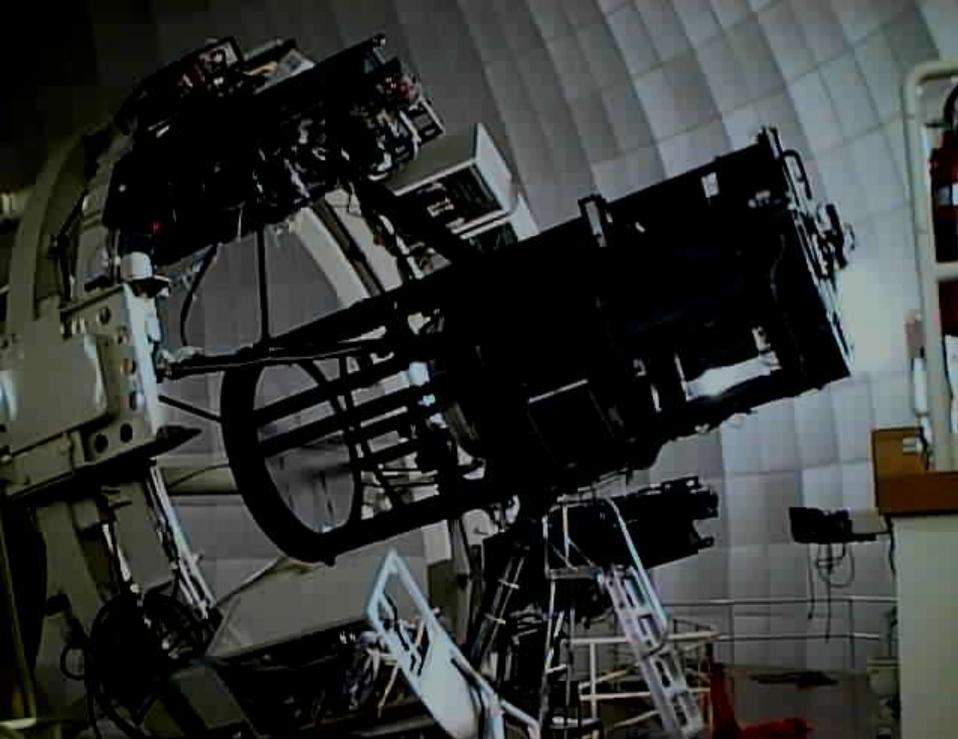
# Journeys in Reproducible Research and Challenges for Institutions, Disciplines and the Individual

**Dr Jonathan Tedds** [jat26@le.ac.uk](mailto:jat26@le.ac.uk) @jtedds

Senior Research Fellow,  
Health & Research Data Informatics, Dept of Health Sciences,  
University of Leicester

PI BRISKit <http://www.brisskit.le.ac.uk>

Director: **RED Informatics Ltd**



Aladin v5.0 \*\*\* BETA VERSION (based on v5.023) \*\*\*

File Edit Image Catalog Overlay View Tool Help

Location 14:02:35.01 +54:12

Object name resolver Ctrl+R  
 Simbad automatic pointer  
 Tooltip on sources  
 Auto-scroll on mouse pan  
 Script Console... PS  
 Macro controller...

VO tools  
 Remote tools  
 Plugins

VO tool controller...

VisIVO 3D visualisation tool [INAF/CINECA].  
 VOFat plotting tool [VO-India]  
 TOPcat tabular data viewer [Starlink/Astrogrid]  
 SPLAT spectral analysis tool [Starlink/PPARC/JAC]

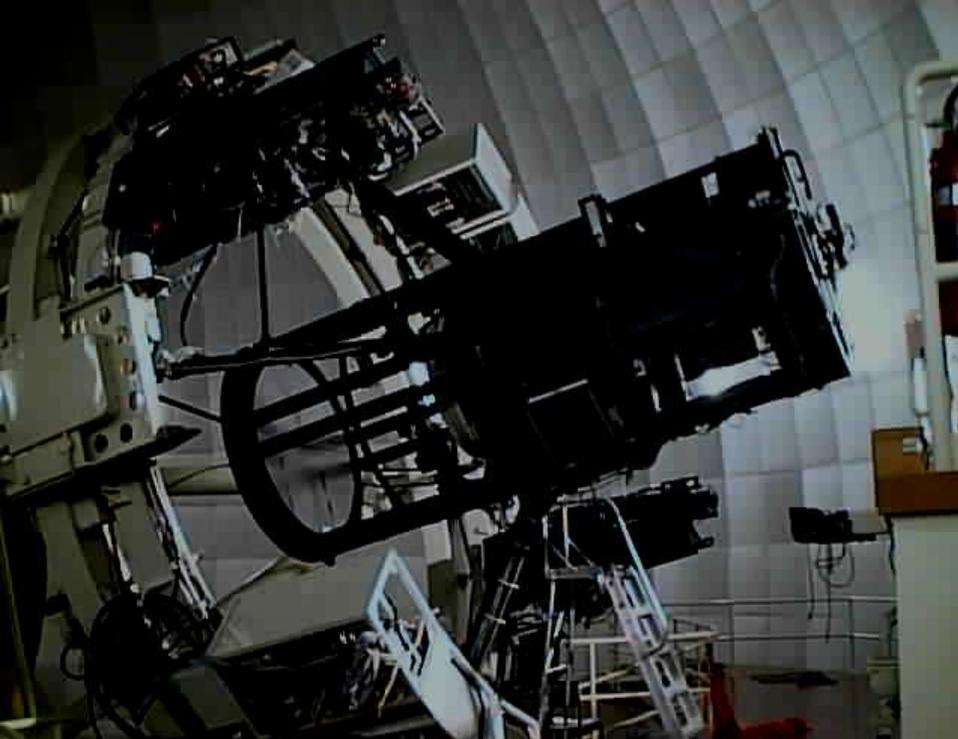
filter  
 NED  
 Simbad  
 DSS1, POSS1

Zoom 1x

MAIN ID	OTYPE	RA	DEC	COO	COO	C	PMRA	PMDEC
<input checked="" type="checkbox"/> IM621	NGC 1969 136	14 02 36.7	+54 21 55	3000	3000	171		
<input checked="" type="checkbox"/> HOOK 179		14 02 36.3	+54 21 46	3000	3000	171		
<input checked="" type="checkbox"/> HOOK 180		14 02 36.8	+54 22 00	3000	3000	171		
<input checked="" type="checkbox"/> NGC 5491		14 02 36.1	+54 21 51	3000	3000	171		
<input checked="" type="checkbox"/> HOOK 176		14 02 35.8	+54 21 54	3000	3000	171		

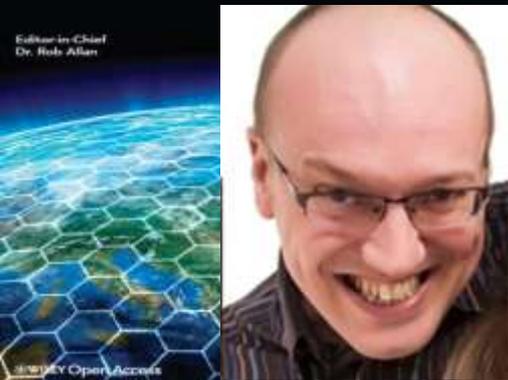
Search

(c)1999-2008 ULP/CHRS - Centre de Données astronomiques de Strasbourg 27 sep / 2020 10:06



Volume 1, Number 1 May 2012

RMet5 Geoscience Data Journal



Open Access



Aladin v5.0 \*\*\* BETA VERSION (based on v5.023) \*\*\*

File Edit Image Catalog Overlay View Tool Help

Location 14:02:35.01 +54:12

Object name resolver Ctrl+R  
 Simbad automatic pointer  
 Tooltip on sources  
 Auto-scroll on mouse pan  
 Script Console... PS  
 Macro controller...

VO tools  
 Remote tools  
 Plugins

VO tool controller...

VoIVO 3D visualisation tool [INAF/CINECA]  
 VOPlot plotting tool [VO-India]  
 TOPcat tabular data viewer [Starlink/Astrogrid]  
 SPLAT spectral analysis tool [Starlink/PPARC/JAC]

filter  
 NED  
 Simbad  
 DSS1, POSS1

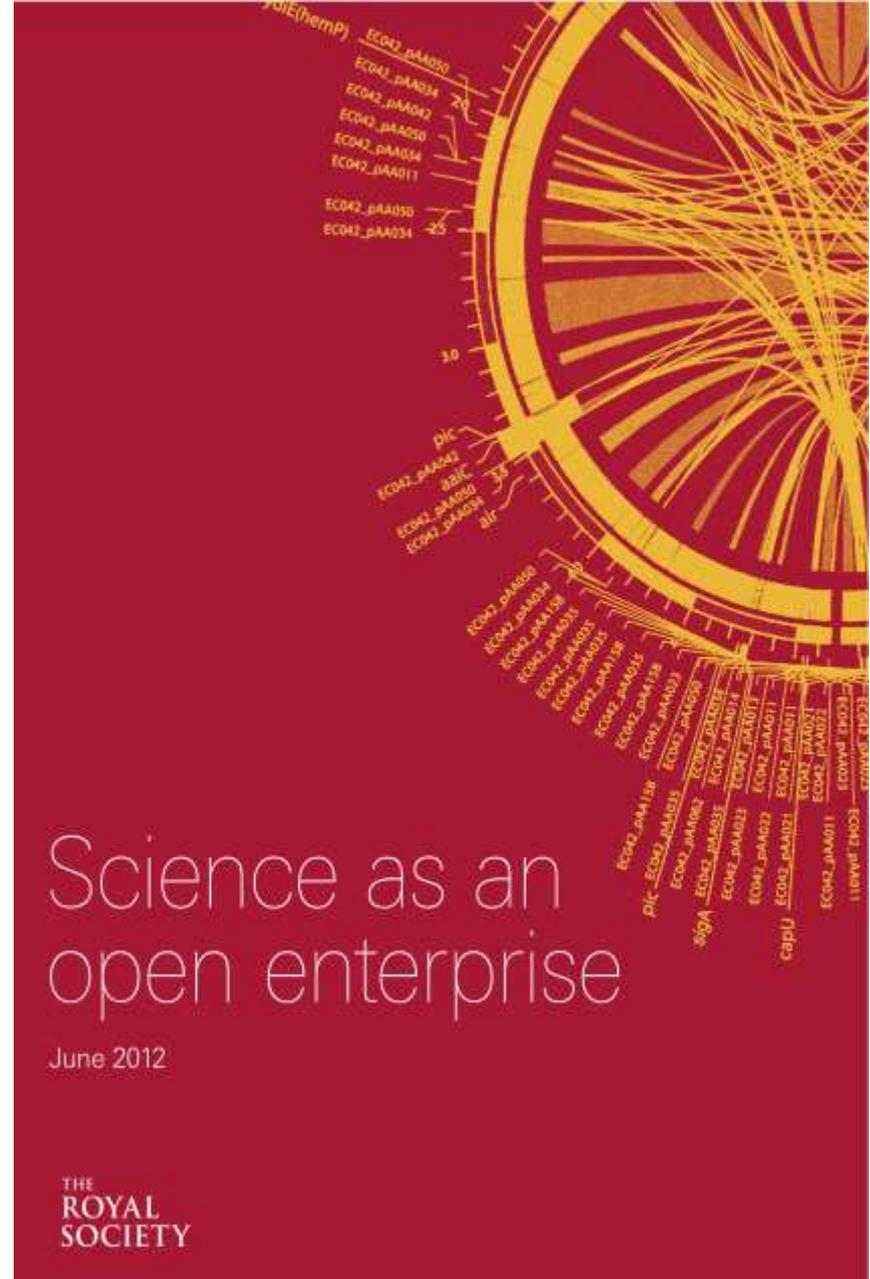
Zoom 1x

MAIN ID	OTYPE	RA	DEC	COD	COD	C	PMRA	PMDEC
<input checked="" type="checkbox"/> IM621 NGC 1969 136	HLI	14 02 36.7	+54 21 55	3000	3000	171		
<input checked="" type="checkbox"/> HOOK 179	HLI	14 02 36.3	+54 21 46	3000	3000	171		
<input checked="" type="checkbox"/> HOOK 180	HLI	14 02 36.5	+54 22 00	3000	3000	171		
<input checked="" type="checkbox"/> NGC 5491	Part	14 02 36.1	+54 21 51	3000	3000	171		
<input checked="" type="checkbox"/> HOOK 178	HLI	14 02 35.8	+54 21 54	3000	3000	171		

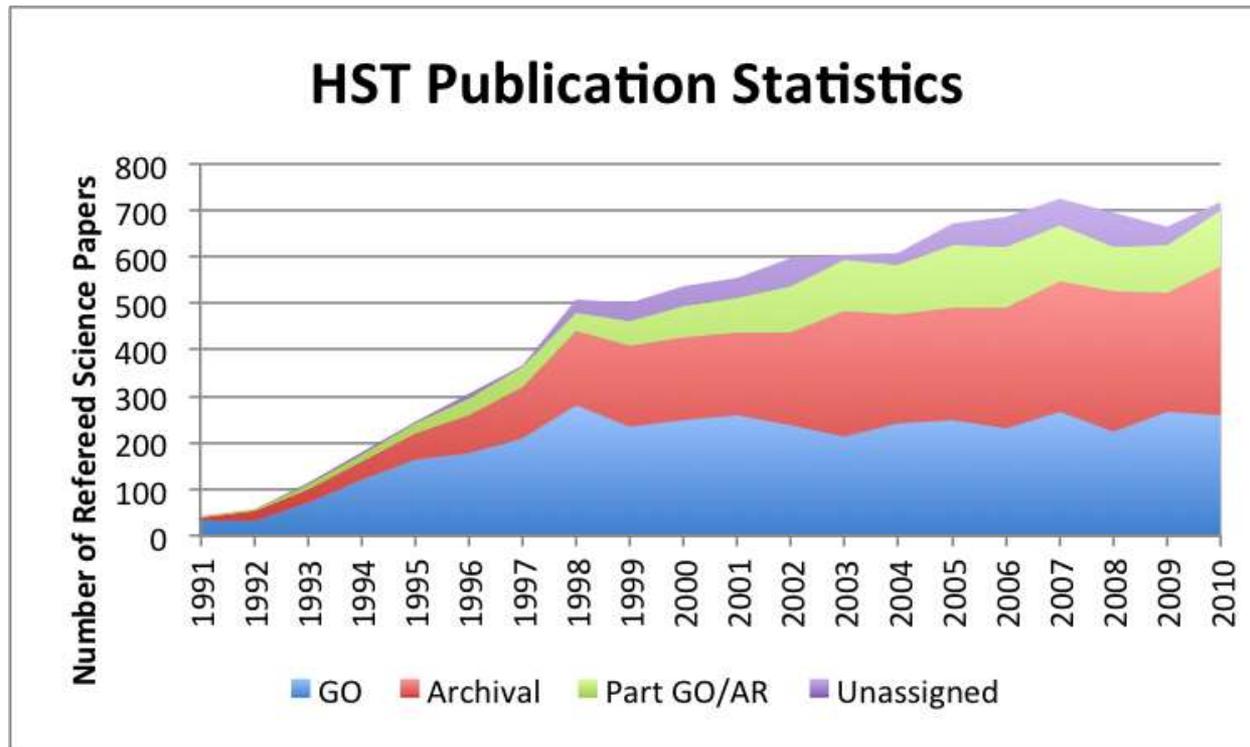
(c)1999-2008 ULP/CHRS - Centre de Données astronomiques de Strasbourg 27 sep / 2012 10:06

# Why open?

- As a first step towards this intelligent openness, data that underpin a journal article should be made concurrently available in an accessible database
- We are now on the brink of an achievable aim: for all science literature to be online, for all of the data to be online and for the two to be interoperable. [p.7]
- Royal Society June 2012, *Science as an Open Enterprise*,  
<http://royalsociety.org/policy/projects/science-public-enterprise/report/>
- Issues linking data to the scientific record:
  - Data persistence
  - Data and metadata quality
  - Attribution and credit for data producers
- Geoffrey Boulton (Edinburgh), Lead author:
  - “Science has been sleepwalking into crisis of replicability...and of the credibility of science”
  - “Publishing articles without making the data available is *scientific malpractice*”



# Data Reuse: asking new questions



## Hubble Space Telescope

- Papers based upon reuse of archived observations now exceed those based on the use described in the original proposal.
  - <http://archive.stsci.edu/hst/bibliography/pubstat.html>
- See also work by Piwowar & Vision re life sciences: “Data reuse and the open data citation advantage”
  - <http://peerj.com/preprints/1/>

# Research and the long tail

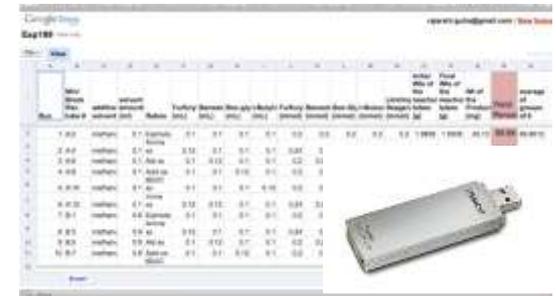
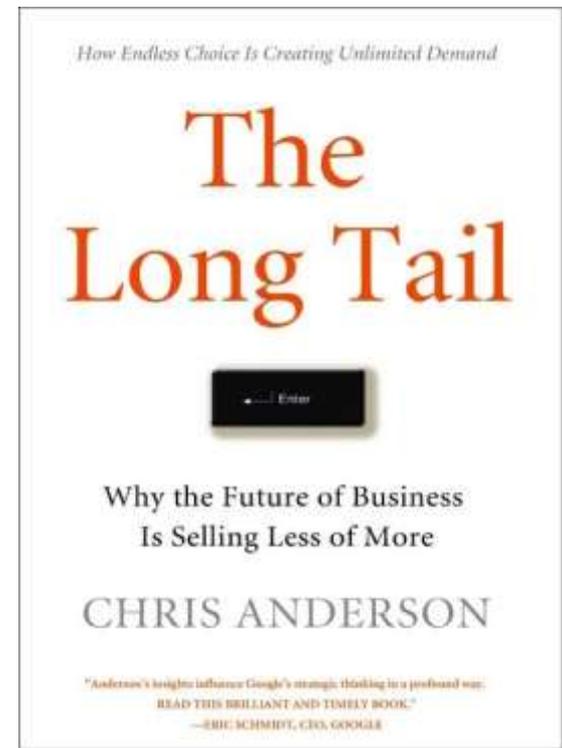


Commons based production  
Public data sets  
Cherry picked results  
Preserved

**GenBank**

**UniProt**

**CATH, SCOP  
(Protein Structure  
Classification)**



**Spreadsheets, Notebooks  
Local, Lost**

# Institutional Perspectives at Leicester

- Some examples...

# 1. HALOGEN (History, Archaeology, Linguistics, Onomastics, GENetics):2009+

*Throwing light on the past through cross-disciplinary databasing*



- ◆ Portable Antiquities Scheme (British Museum)
- ◆ Place-names (Nottingham)
- ◆ Surnames
- ◆ Genetics
- ◆ IT hosting and GIS
- ◆ Best practice: #JISCMRD, UKRDS, DCC, international

<http://halogen.le.ac.uk>



University of  
**Leicester**

**welcome**trust



## Key to English Place-Names

Data provided by the Institute for Name-Studies at the University of Nottingham.



County boundaries - © Great Britain Historical GIS Project,  
University of Portsmouth.

### Search by county:

Click on the counties on the map, or select from the list. Hold Ctrl to select multiple counties from the list.

#### ALL COUNTIES

-----  
 Bedfordshire  
 Berkshire  
 Buckinghamshire  
 Cambridgeshire  
 Cheshire  
 Cornwall

### Search by language:

(Hold Ctrl to select multiple languages.)

#### ALL LANGUAGES

-----  
 Celtic  
 French  
 Old English  
 Old Norse  
 Latin  
 Anglian

### Search by element:

Hover your mouse over the elements for their definition, or look at the list of [valid elements](#). Hold Ctrl to select multiple elements from the list.

#### ALL ELEMENTS

-----  
 -aco-  
 -an  
 -binde  
 -dwestle  
 -ek  
 -ell

### Search by place name:

Exact match

# Halogen as template for research data management

#jiscmrd

- Requirements Analysis - must be iterative!
- *Data Management Plan - use DMPonline (UK Digital Curation Centre)*
- Scalable research data management infrastructure
  - pilot phase to nationally available resource
  - LAMP stack IT infrastructure: host research database
- A model for the long term delivery of a data management service within the institution including
  - support, maintenance, governance & charging policies
  - Include researchers, IT services, research support office, library services etc.

# BENEFITS

- *New research opportunities*
  - **Cross database work - seed new research samples**
- *Scholarly communication/access to national resources*
  - Key to English Place Names (Nottingham)
  - Portable Antiquities Scheme (British Museum)
- *Verification, re-purposing, re-use of data*
  - Cleaning & enhancing private research datasets for reuse & correlation
  - No re-creation of data
  - Increased transparency
  - excellent training for best practice in research data management
- *Increasing research productivity*
  - Build in cleaning, annotation, enhancement into normal research workflows
  - research datasets may immediately be reusable and interoperable
- *Impact & Knowledge Transfer*
  - Reuse IT infrastructure
- *Increasing skills base of researchers/students/staff*

# CHALLENGES

- ***interdisciplinary research database***
  - ingest each input dataset in form such that sufficient information is carried forward to enable interoperation
  - ***Cultural differences!***
- ***versioning & provenance*** for input datasets
- ***which software tools, infrastructure , Query interface?***
  - suitable for multi disciplinary researchers
- ***Requirements upon the institution*** for sustaining the research assets & skills
- ***Requirements upon the researchers***
  - Annotating
  - Refreshing
  - Maintenance of datasets

*Reward = Leverhulme Trust funding £1.3m! Helped enable...*

# The Impact of Diasporas on The Making of Britain



*Evidence*  
*Memories*  
*Inventions*





Home » About The Centre » An incredible discovery

## An incredible discovery



### BOOK TICKETS



To make sure you can visit the Centre at a time that suits you, we strongly recommend you book tickets in advance. It's quick and easy to book online

### EXPERIENCE THE LEGEND



# Top Tip for institutions: how to get researchers' attention?

Modify Costing No. 9 (Application Version)

Research Data Management

**IMPORTANT:** The University and funding bodies require that researchers demonstrate that they have planned for the management of research data over the project life cycle. Please indicate if:

1. Your research may use or generate research data that is sensitive or confidential (including Commercial in Confidence)?  Yes

2. The project will require advice in planning and costing your IT and research data management requirements (including secure and reliable storage) either before, during or after completion?  Yes

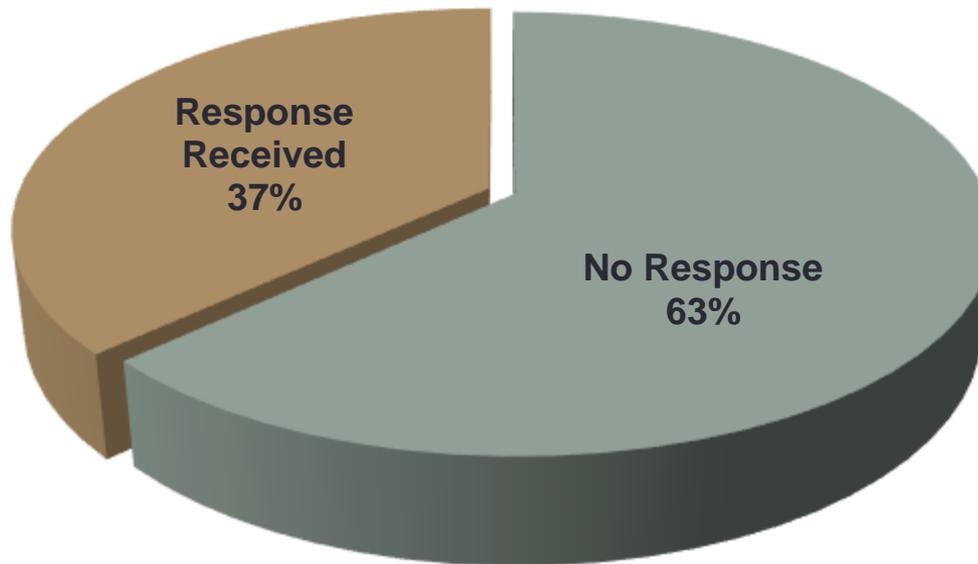
Responding 'Yes' to the above questions will be logged with IT Services for the appropriate team to make contact.  
Please tick this box if you wish to hide the title of the research project in this notification  Hide Title

Please contact the [IT Academic and Research Liaison team](#) for advice in good time before submission.

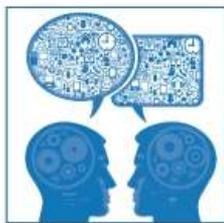
Continue  Cancel

# Research costing - only part of the answer

**Researcher Responses to Contacts Made**



## Find & Share

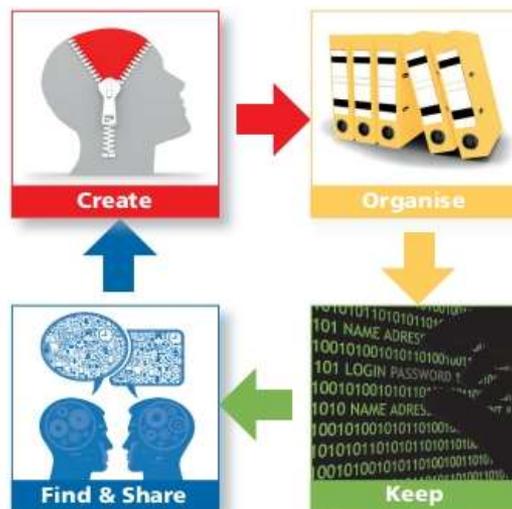


### Do you know how to...

- find existing information resources related to your research?**  
Where can you find research data that you can repurpose or combine with your own to produce new research?
- share data with your collaborators securely and effectively?**  
Whether building a collaborative proposal, generating results for others to comment on or sharing the final outputs of your research – how will you link with your colleagues ahead of wider sharing?
- deposit your research data and outputs in an open repository?**  
Is there an appropriate disciplinary or institutional repository and what do you need to do to deposit your research output? Plan ahead to avoid refactoring...
- publish your research, and get it cited as well?**  
Institutions and data centres must make research data reusable to others while providing credit to the researchers who did the work. Your future career could depend on it!

Chances are you could use some helpful pointers in all of these!

To find information, support, advice and training, as well as links to external resources, go to [www.le.ac.uk/researchdata](http://www.le.ac.uk/researchdata)



email: [researchdata@le.ac.uk](mailto:researchdata@le.ac.uk)

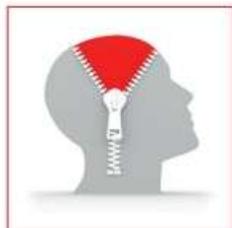


## What would you do if you lost your research data tomorrow?

*Take the research data health check... and find help to secure, share and exploit your valuable research.*



## Create



### Have you...

- fully understood your research funders' data management requirements?**

Government and funders require that publicly funded research is made available for reuse – are you up to date with their latest policies? Your future funding might depend on it!

- written a data management plan?**

Your funder may already require this but build it in from the proposal stage to avoid headaches in the future.

- gained ethics approval/consent?**

Writing a data management plan will aid planning and help you to navigate ethics and governance requirements.

- protected your intellectual property?**

Leaving intellectual property considerations for a rainy day could lose you appropriate credit, damaging career prospects and perhaps your financial future health!

## Organise



### Are your research files and data...

- clearly described, in terms of content (using standard metadata)?**

Ask yourself honestly: is there a danger of data being lost? Will you be able to remember how you generated your data, and will you or anyone else be able to find it in the future when you wish to reuse and share?

- clearly labelled with versions and dates?**

How will you remember which was **the** definitive version and which dataset was used in producing a given research outcome?

- logically structured and named?**

Once you've remembered **how** you generated data, can you still **find** the relevant files?

- future-proofed against broken links, using persistent identifiers?**

The persistent identification of digital resources can play a vital role in enabling their accessibility and re-usability over time using recommended data standards.

## Keep



### Do you know...

- how to restrict access to your research data to the right people?**

Have you consulted with university or data centre experts so that only the right people have access to your research?

- which data to keep and which data to discard?**

Managing research data effectively means being selective: **which** data to discard and **when** as well as **what** to keep and for **how long**?

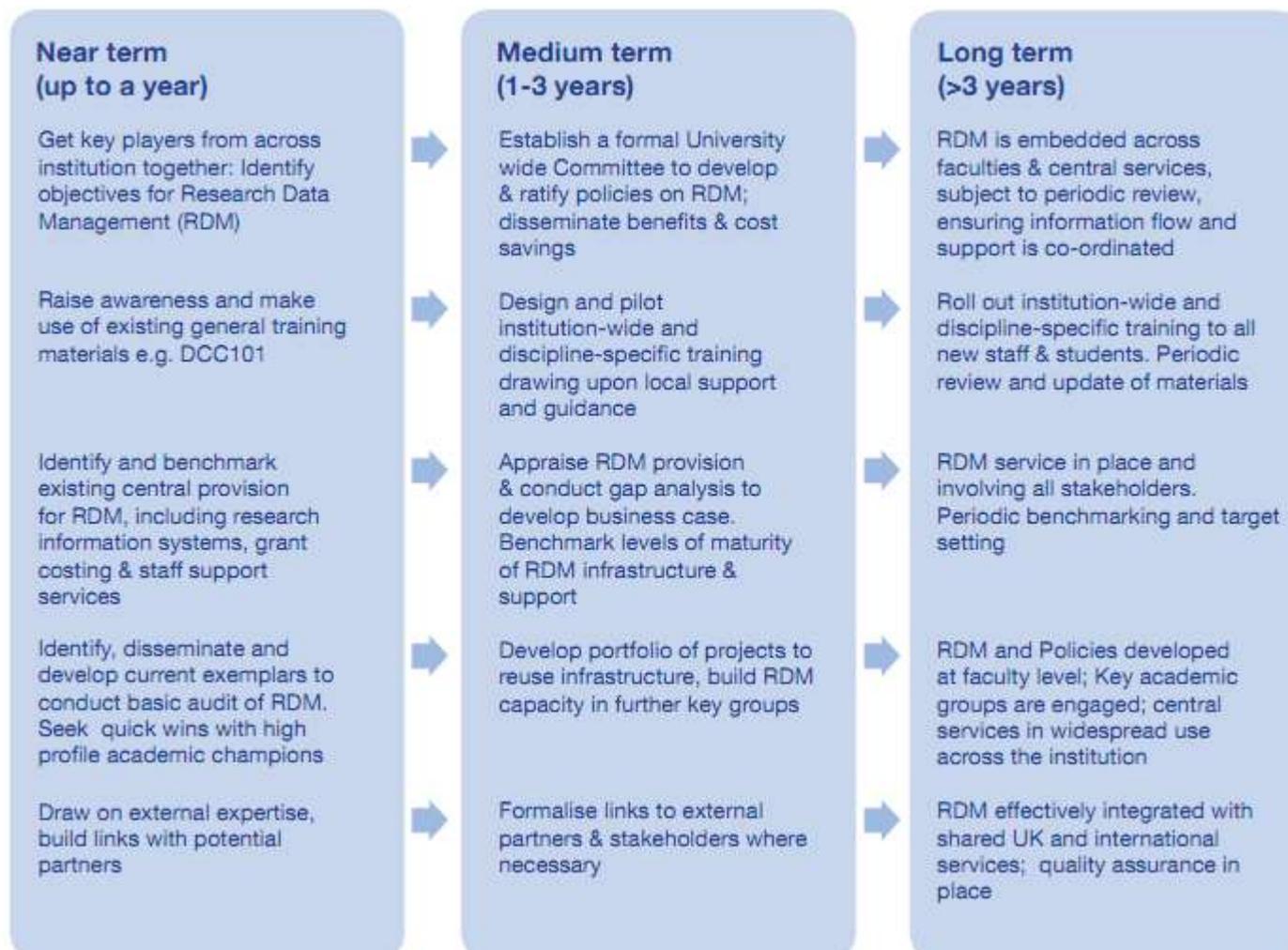
- how securely your data is stored?**

What happens if your storage media fail? How resilient is it? Could it get left on the train and could somebody else misuse it?

- how your data is backed-up?**

Have you made use of university and/or external resources to back up data so that you have multiple copies in case of loss or theft?

# Suggested timeline for implementing institutional research data management



From Whyte & Tedds (2011), DCC Briefing

<http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm>

# Challenges for researchers in institutions

- **Rise to scientific and research challenge**

- Not just a management challenge
- Responsibility for the knowledge you create
- Personal benefits as well as funder compliance

- **Library**

- curation of data and publications
- active support from data scientists
- from centralised to dispersed support
  - Expert centres essential!

- **IT Service**

- Provide research data platforms for researchers:
  - Active storage
  - Enable collaboration
  - Connect to preservation services through Library

# Enabling Research Data Management

- Active Data Management Planning
  - built in at proposal stage
    - Local institutional tweaks of funder and local templates
  - Implemented and evolved in project
    - Data Management Plan as a live, evolving object
    - Annotate data on the fly - lab notebook approach
  - Curated & preserved using permanent identifiers
    - Appropriate repository and data collection descriptors

# It's a long road....

What do researchers need to make this all possible?

- Incentives - citations, promotion, support - long way to go
- Institutional and funder policy framework - mostly there now?
- Appropriate discipline specific community centres of expertise - rare, mostly limited to big science niches or very broad but poorly sustained
- Institutional support services for the basics - pilots to date
- Software tools that are open and can be adapted - on the way

# But that's not all...

- What about the software underpinning data driven research?
- If we're going to publish as open data:
  - How do we help researchers to store, annotate and discover the datasets they create?
  - How do you sustain and reuse that?

# How important is software to research?

- **SSI survey (2014)**
  - **92%** of academics **use** research software
  - **69%** say that their research **would not be practical** without it
  - **56% develop** their own software
    - **21%** of those have **no training** in software development

<https://www.software.ac.uk/blog/2014-12-04-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers>



# Stakeholders & Motivation

- The further from the creation of the code, the greater the interest in identifying and preserving it is.
  - **Research software engineers:**
    - “Good software management practice is all that is needed”
    - We suspect those who need to reuse code may not agree ...
  - **Computational scientists who write code:**
    - Haven’t thought about it but acknowledgement/credit and reproducibility are good in theory
  - **Digital Preservation experts:**
    - Very interested as they know they will have to do it



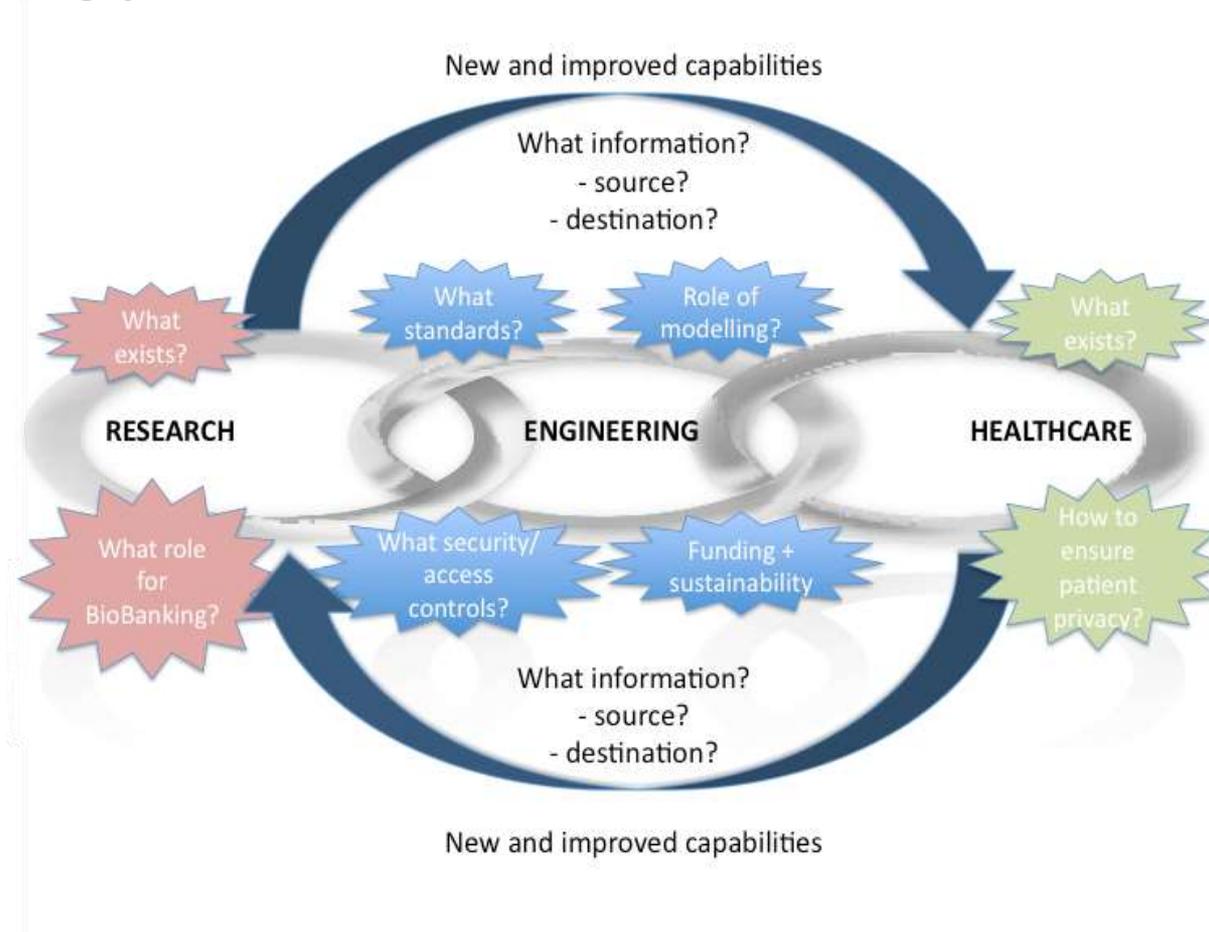
## **2. Biomedical Research Software as a Service**

A vision for cloud-based open source research applications  
**#BRISSKit**

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

# context:

The I4Health goal of applying knowledge engineering to close the 'ICT gap' between research and healthcare (Beck, T. et al 2012)



[www.brisskit.le.ac.uk](http://www.brisskit.le.ac.uk)

Email: [brisskit@le.ac.uk](mailto:brisskit@le.ac.uk)



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**

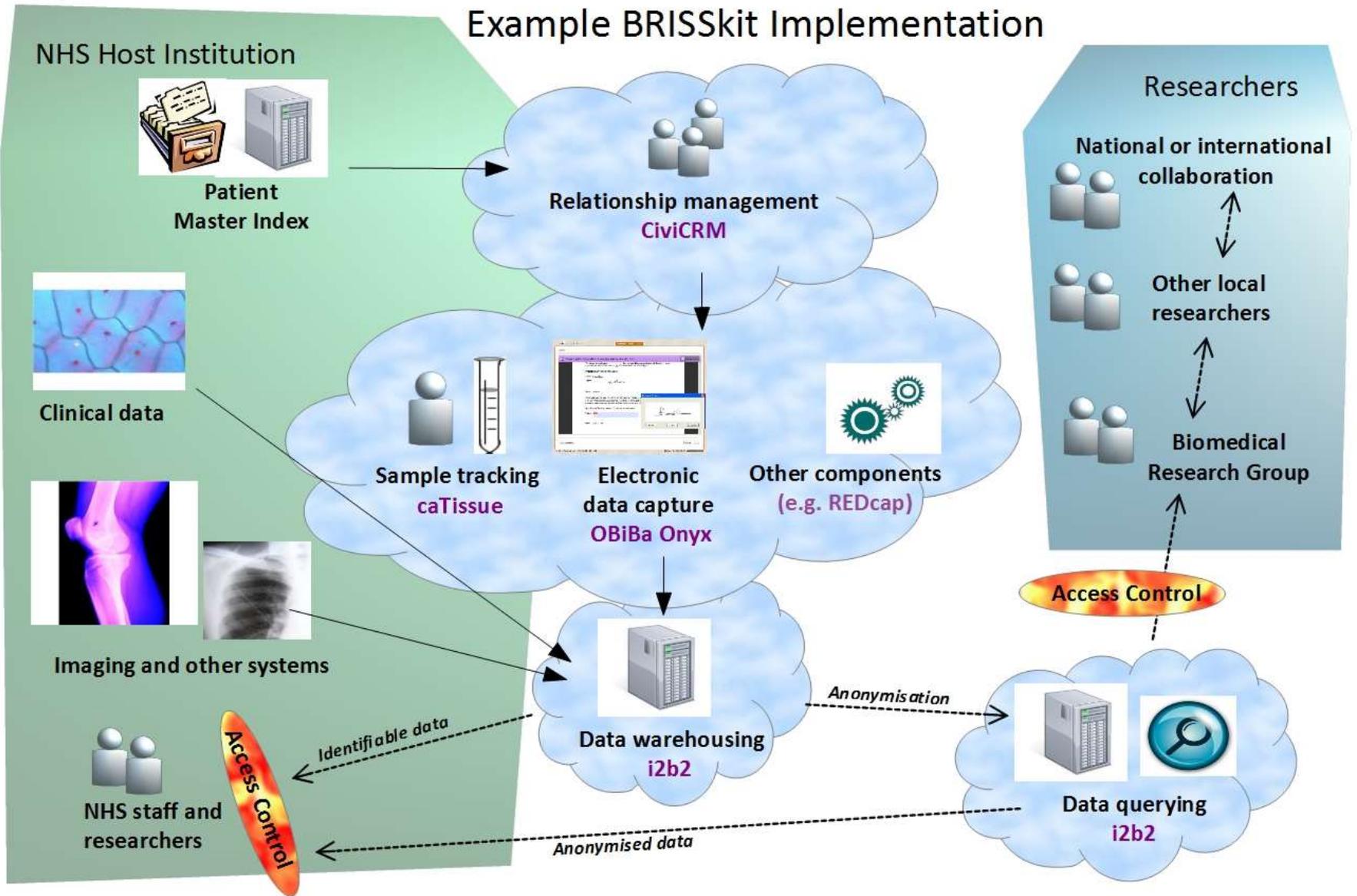
Jisc

University Hospitals of Leicester

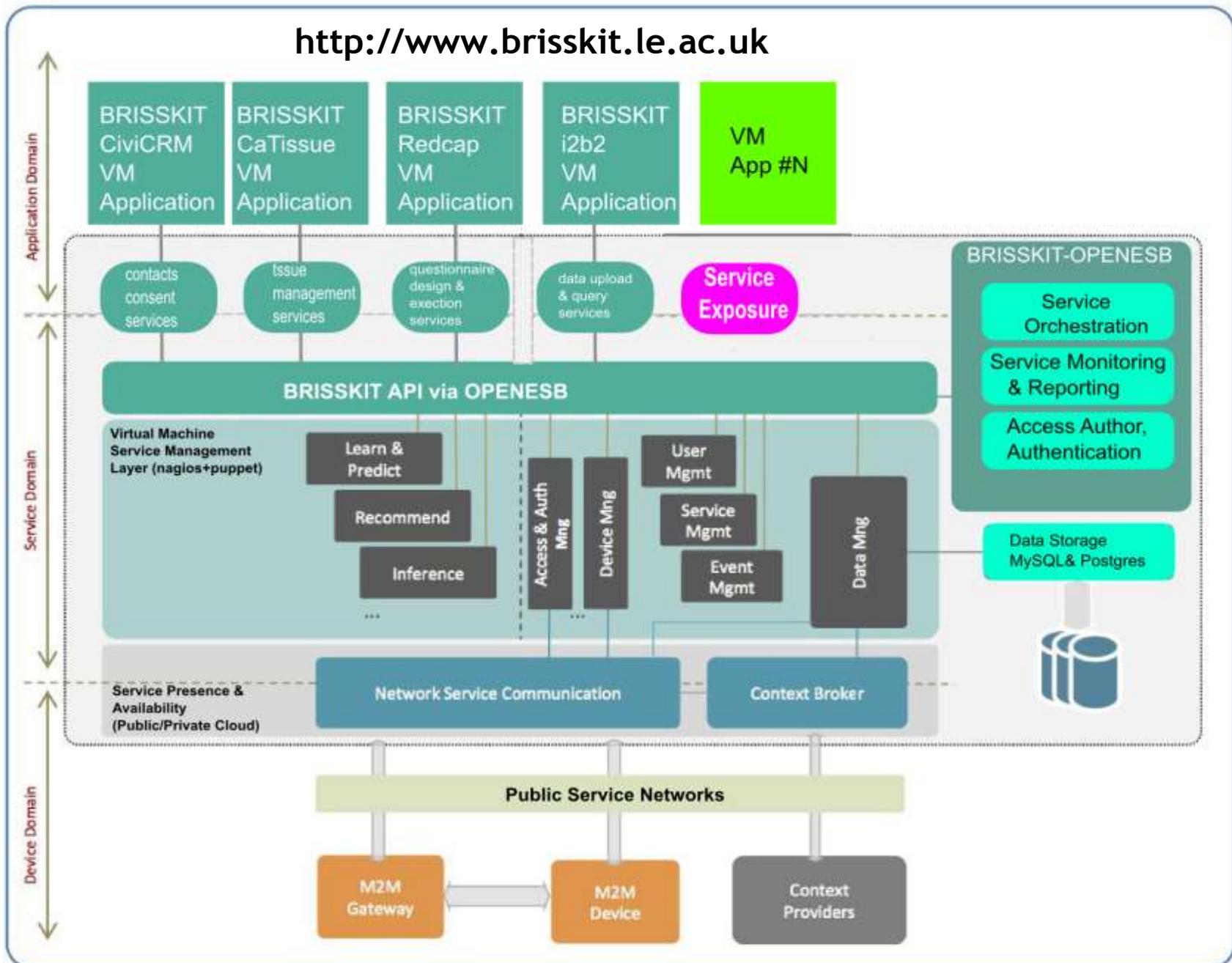


NHS Trust

# Example BRISKit Implementation



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



# BRISKit CiviCRM: patient cohort management

The screenshot displays the BRISKit CiviCRM interface. The top navigation bar includes links for Home, Search, Contacts, Recruitments, Reports, Administer, and Help. The main content area is titled 'CiviRecruitment Dashboard' and features a 'Summary of All Recruitments' table and a table of 'All Recruitments With Upcoming Activities'.

**Summary of All Recruitments**

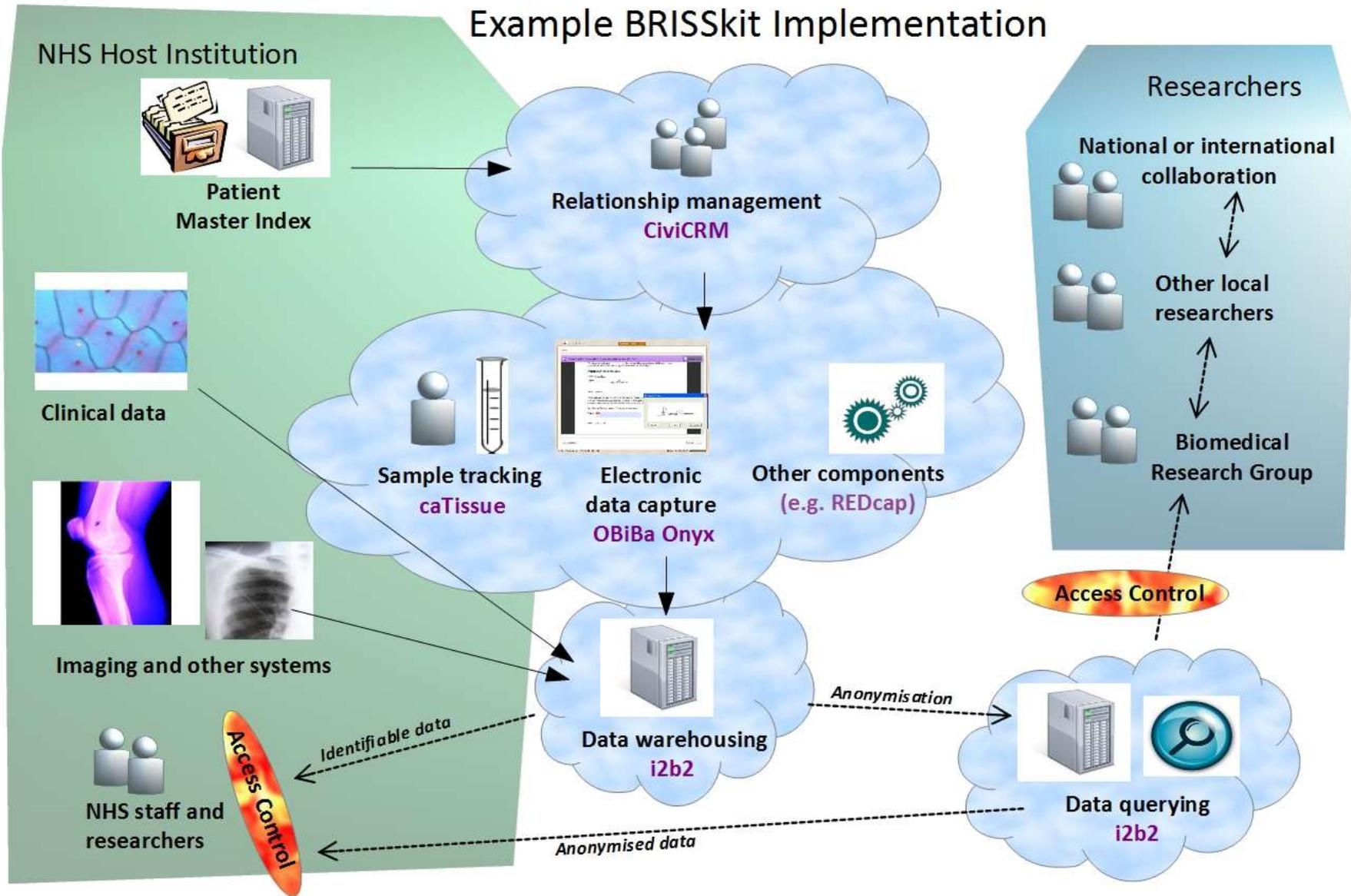
	Ongoing	Resolved	Urgent	Self-withdrawn	Completed	Screen fail	In study	Investigator withdrawn	Lost to followup	Other
Housing Support	0	0	0	0	0	0	0	0	0	0
Adult Day Care Referral	0	0	0	0	0	0	0	0	0	0
RespiratoryResearch	39	0	0	0	1	0	0	0	0	0

**All Recruitments With Upcoming Activities**

Client	Subject	Status	Type	My Role	Recruitment Manager	Next Sched.
BLOMPENHEFFLE, JANICE 01509 377665 Recruitment ID: 1	Automatically added to RespiratoryResearch study	Ongoing	RespiratoryResearch	---	NDRI Leicester Respiratory Biomedical Research Unit	Organise appointment November 8th, 2013 12:00 AM
HUGGLESPRINT 0114 2888333 Recruitment ID: 2	Automatically added to RespiratoryResearch study	Ongoing	RespiratoryResearch	---	NDRI Leicester Respiratory Biomedical Research Unit	Consent to be added to recruitment DB November 8th, 2013 12:00 AM

- Manages studies: enables end-to-end contact management for volunteers and research participants
- track approaches, contact, responses, recruitment, exclusions
- In use for 22k recruits in 14 studies (Cardiovascular); new development for NIHR Lifestyle BRU, 100k Genomes

# Example BRISKit Implementation



# BRISKit OpenSpecimen: sample management

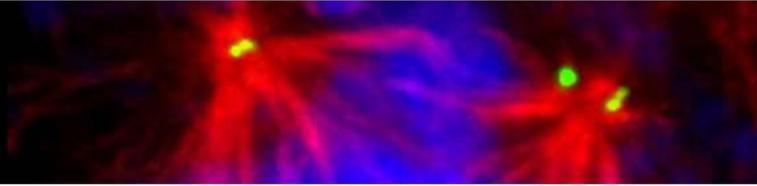
The screenshot displays the OpenSpecimen 1.0 web application. The browser address bar shows the URL: `catissue31.8000/OpenSpecimen/Login.do?loginName=cj.issa@gmail.com&password=Login1234567`. The application header includes the OpenSpecimen logo and navigation links for Administration, Biobanking, Search, and Admin Admin. The main content area is divided into several sections:

- Collection Protocol:** A dropdown menu showing "c1short" with a "Watch Tutorial" link.
- Participant:** A dropdown menu showing "Issa,(1\_1)" with "Register Participant" and "View Participant" buttons.
- Specimen Tree:** A tree view showing a hierarchy of specimens. The selected specimen is "168\_6n(Fixed Cell Block)", which is a child of "1687(Fixed Cell Block)".
- Specimen Details:** A form with the following fields:
  - Parent Label: 1687j
  - Label: 168\_6n
  - Class: Cell
  - Tissue Site: Not Specified
  - Pathological Status: Not Specified
  - Initial Quantity: 1.0 cell count
  - Is Available?:
  - Collection Status: Collected
  - Storage Position: Virtually Located
  - Comments: (empty text area)
  - External Identifier(s): [Add New](#)
  - Biohazard(s): [Add New](#)
  - Lineage: Aliquot
  - Barcode: 1\_1999888
  - Type: Fixed Cell Block
  - Tissue Side: Not Specified
  - Created On: (empty date field)
  - Concentration: (empty field)
  - Available Quantity: 1.0 cell count
  - Activity Status: Active
- Create Child Specimen(s):** Radio buttons for "None", "Aliquot", "Derivative", and "As per Protocol". A "Print Labels" checkbox and "Submit" button are also present.

- Holds data on primary, derived and aliquot specimen, including linear and 2d barcodes
- Storage inventory, order tracking e.g. 100,000+ samples NIHR UHL Cardiovascular BRU in 5+ yrs; Cancer Biobank - next slide



LEICESTER  
CENTRE



Cancer Research UK

About Us

Research

Centre Network Accelerator

Trials

Facilities

→ The Central Research Facility

→ Structural Biology

→ **Cancer Biobank**

– Samples

– Contacts

→ Bioinformatics

Study and Jobs

Community

Staff

Contact

## Leicester Cancer Research Biobank

Cancer research is currently in an era in which thousands of genes, RNA molecules and proteins can be analysed in a single experiment, meaning that large numbers of potential molecular markers for cancer are being identified. To validate and translate these biomarkers large series of tissue and other biosamples are required.

The **Cancer Research UK Leicester Centre** and University Cancer Theme which oversees the cancer research strategy at Leicester, recognises that a joined up approach to cancer sample collection and storage is needed. The University of Leicester Cancer Research Biobank is intended to meet this need and to facilitate high quality research.

The Biobank will support projects from all sections of the cancer research community within Leicester, while remaining open to provision of tissues to organisations outside Leicester.

Researchers from academic, charitable and commercial organisations will be able to apply for access to banked tissue samples.



The Leicester Cancer Research Biobank is a member of, and accredited by, the **National Cancer Research Institute - Confederation of Cancer Biobanks**.

The Leicester Cancer Research Biobank also has full ethical approval, bestowed by NRES Committee East Midlands Derby (Ref: 13/EM/0196; IRAS 103613) and is operated under Human Tissue Act Licence 12384.

- **Open Specimen - Sample tracking for LCCB, Tracer X, Mesobank**

Share this page:  Like  0  Tweet  G+  0  Share

# BRISKit RedCap: survey management

The screenshot shows the REDCap web interface for a project named 'Questionnaire Simple'. The interface is divided into a left sidebar and a main content area. The sidebar contains navigation links for 'My Projects', 'Project Home', 'Project Setup', 'Data Collection', 'Invite Participants', 'Add or View Survey Responses', and 'Applications'. The main content area has a title 'Questionnaire Simple' and three tabs: 'Project Setup', 'Online Designer', and 'Upload Data Dictionary'. Below the tabs, there is a 'VIDEO: How to use this page' link and a paragraph explaining how to build and customize data collection instruments. A 'Preview instrument' button is visible. The main content area displays two survey questions with radio button options:

Variable: aunits  
How many units of alcohol do you have on a typical day when you are drinking?  
 1-2  
 3-4  
 5-6  
 7-8  
 9+

Variable: apy  
How often did you have a drink containing alcohol in the past year?  
 Monthly or less  
 4 times per month  
 4 times per week  
 4 or more times per week

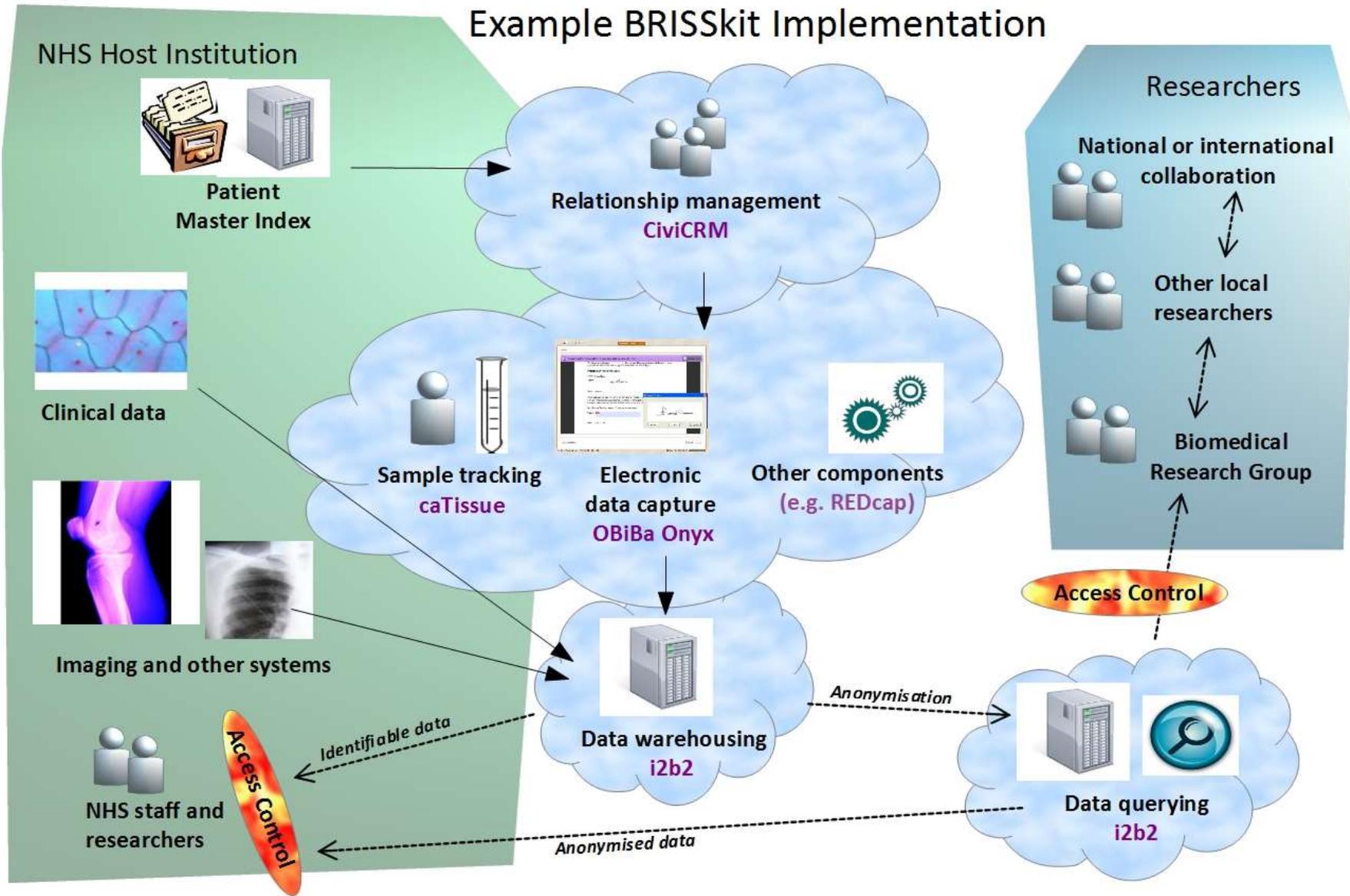
Web-based, secure questionnaire data entry by research or nursing staff

E.g. used for all patient recruits in NIHR UHL Respiratory Biomedical Research Unit - mobile computing on wards and outpatient clinic

- Exceed Study (Tobin)
- Royal Derby NHS

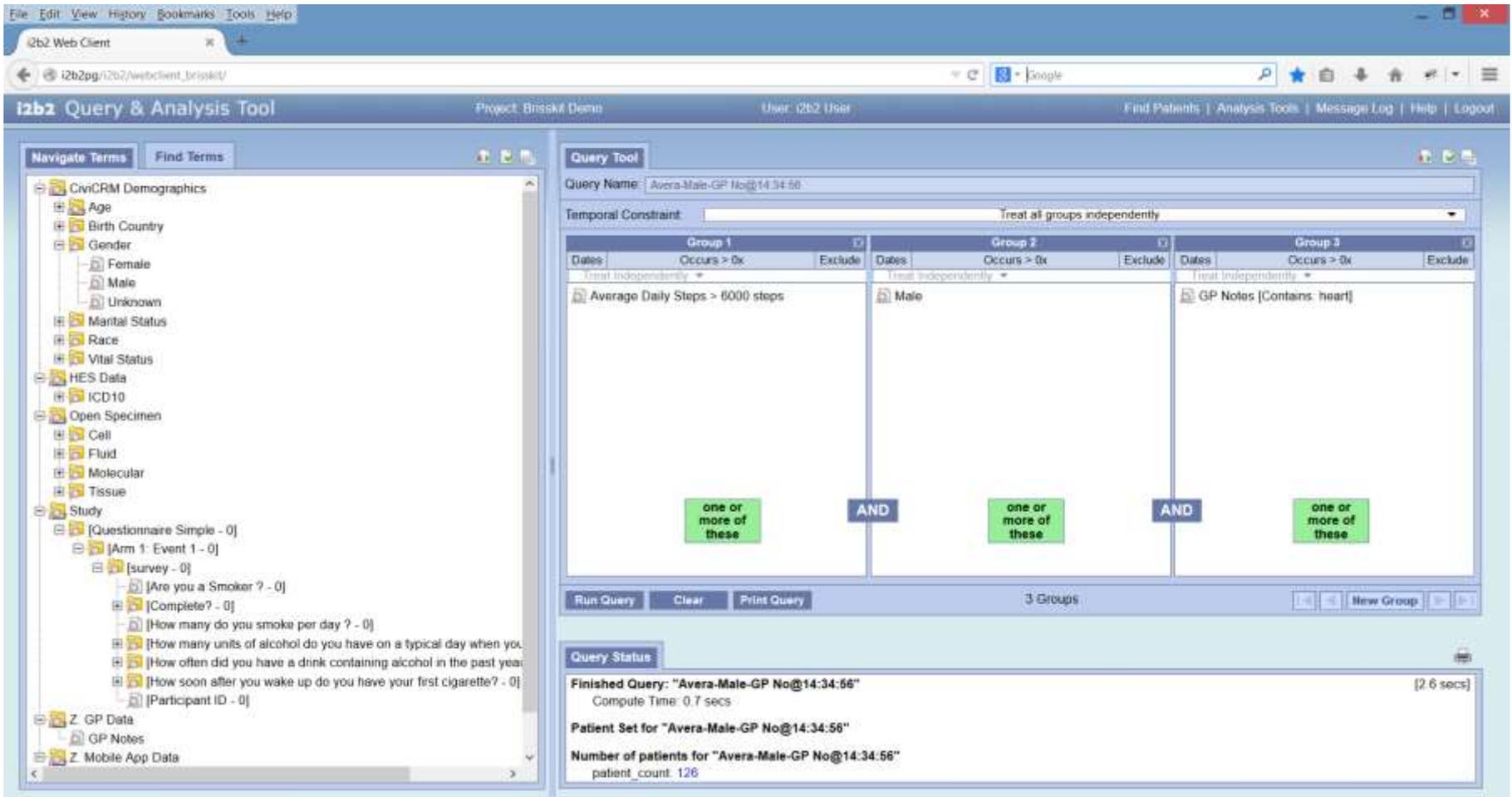
<http://www.brisikit.le.ac.uk>

# Example BRISKit Implementation



<http://www.briskit.le.ac.uk>

# BRISKit i2b2: data warehousing & querying



The screenshot displays the BRISKit i2b2 Query & Analysis Tool interface. The interface is divided into several sections:

- Navigation Tree (Left):** A hierarchical tree structure showing data sources. Key categories include:
  - CiviCRM Demographics
    - Age
    - Birth Country
    - Gender
      - Female
      - Male
      - Unknown
    - Marital Status
    - Race
    - Vital Status
  - HES Data
    - ICD10
  - Open Specimen
    - Cell
    - Fluid
    - Molecular
    - Tissue
  - Study
    - [Questionnaire Simple - 0]
      - [Arm 1: Event 1 - 0]
        - [survey - 0]
          - [Are you a Smoker? - 0]
            - [Complete? - 0]
            - [How many do you smoke per day? - 0]
            - [How many units of alcohol do you have on a typical day when you...]
            - [How often did you have a drink containing alcohol in the past year...]
            - [How soon after you wake up do you have your first cigarette? - 0]
            - [Participant ID - 0]
    - Z. GP Data
      - GP Notes
    - Z. Mobile App Data

- Query Tool (Center):** A panel for constructing queries. It shows:
- Query Name: Avera-Male-GP No@14:34:56
- Temporal Constraint: Treat all groups independently
- Three groups of criteria:
  - Group 1: Average Daily Steps > 6000 steps
  - Group 2: Male
  - Group 3: GP Notes [Contains: heart]
- Operators: AND between groups, and 'one or more of these' for each group.
- Buttons: Run Query, Clear, Print Query, New Group.
- Query Status (Bottom):** A panel showing the results of the query:
- Finished Query: "Avera-Male-GP No@14:34:56" [2.6 secs]
- Compute Time: 0.7 secs
- Patient Set for "Avera-Male-GP No@14:34:56"
- Number of patients for "Avera-Male-GP No@14:34:56": patient\_count. 126

Data from multiple data sources combined into multiple ontologies for flexible and sophisticated searching, cohort discovery and research

# Research: the semantic bridge

Survey data

Records participant consent, questionnaire data and primary specimen IDs

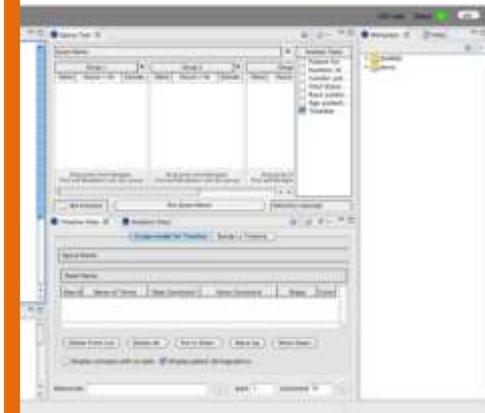


Bio-ontology!

- [-] Classes
  - [-] BrissKitQuestionnaire
    - [+] ← RiskFactorAlcohol
    - [+] ← RiskFactorCigars
    - [+] ← RiskFactorCigs
    - [+] ← RiskFactorDietExercise
    - [+] ← RiskFactorFamilyHistory
    - [+] ← RiskFactorOtherTobacco
    - [+] ← RiskFactorSocioeconomic
    - [+] ← RiskFactorTobacco
  - [-] HPO
    - [+] ← All
    - [+] ← Phenotypic abnormality
      - [+] ← Abnormality of the cardiovascular system
      - [+] ← Abnormality of the hematopoietic system
  - [-] ICD10
    - [+] ← Diseases of the circulatory system
    - [+] ← Diseases of the skin and subcutaneous tissue
    - [+] ← Factors influencing health status and contact with health services
      - [+] ← Persons encountering health services in other circumstances
  - [-] LOINC
    - [+] ← LOINCCLASSTYPES
      - [+] ← Clinical Class
        - [+] ← Functional status (e.g. Glasgow)
        - [+] ← Gastrointestinal endoscopy
        - [+] ← Medical Devices
  - [-] SNOMED-CT
    - [+] ← Clinical finding
    - [+] ← Functional finding
    - [+] ← Hepatorrhaphy
    - [+] ← Observable entity
    - [+] ← Organism
    - [+] ← Social and personal history finding
    - [+] ← Substance
    - [+] ← Tobacco smoking behavior - finding
    - [+] ← Tobacco use and exposure - finding

2 data query

port selection and querying



# BRISKit Information Governance & Security Management Work Stream

1. **Information Governance Toolkit** - analysis of Department of Health (DoH/NHS) IGT requirements vs. BRISKit organisation/project and services/tools
  - a) Hosted Secondary Use Team/project ([Hosted IGT](#))
  - b) Acute Trust ([Acute Trust IGT](#))
  
2. **IG Training Tool** (NHS – University is registered)
  - UoL College of Medical Biological & Psychological Sciences wide registration
  
3. **Pseudonymisation** requirements
  
4. **Data Management Plan**
  
5. **IT Security & standards** – Penetration Testing & Security Testing
  
6. **Other NHS Standards/Requirements:**
  - Care Records Guarantee
  - NHS Constitution
  - NHS Records Management
  - Patient Safety DSCN 14/2009, 18/2009

***NB – Partnerships available via MS Azure for Research, Aimes, UK Cloud (Skyscape) for secure IGT compliant DB hosting***

# Code Release

- **code base** freely available to the community via Github repositories
  - 95+ repositories released
  - licences
  - Fully documented
- set of **documentation** and explanatory materials to support further take up and use
  - puppet management, maven build manager and provisioning of virtual machines
  - integrations between the applications
- final release July 2016
- Docker implemented

## Apperta Foundation Challenge

Our challenge is now closed, join us at EHI Live in November 3rd/4th where the results have been announced.

<http://www.ehilive.co.uk/>

The Apperta Foundation for Health and Care announced a Code4Health challenge at the Health & Care Expo 2-3 September in Manchester, challenging interested parties to submit ideas which utilises existing or new open source resources to provide a solution they feel is required across health and social care.

Submissions were assessed, all are now shortlisted and are asked to create a Code4Health Community (if one doesn't exist) and will be provided with guidance and resources via the Apperta Foundation to create and develop the solutions further.

### Apperta Foundation Challenge Submissions

#### NET

Bolton CAMHS

Mobile Safety Plan for Children and Young People with Mental Health Difficulties

#### Care Home Dementia Diagnosis

Yorkshire & The Humber Strategic Clinical Network

Making a diagnosis of established dementia in a care home setting

#### 'What's the Plan?'

'What's the Plan?' Code4Health Community

The Open Care Plan

#### National Cataract Audit Platform

Moorfields Eye Hospital/Cardiff University

Making a diagnosis of established dementia in a care home setting

#### Appetite

Guys & St Thomas NHS Trust

Web based personalised dietary tool for patients with cancer

#### Pathways

Guys & St Thomas NHS Trust

Web based personalised information delivery

#### BRISKIT

BRISKit

Integrating specimen data for analytics using open source tools

#### Safety in Your Pocket

Bolton CAMHS

Mobile Mental Health application straight out of NHS Hack Day Manchester

#### Electronic Discharge Summaries

HSCIC

Electronic Discharge Summaries from Acute to Primary Care

# 100k Genomes Key Activities

- *Recruitment, recruitment, recruitment!*
- Refining sampling process & flows
- Phenotyping & Pedigrees
- EHR Integration
- Clinical data analysis

# 100k Genomes in East of England GMC

- University Hospitals Leicester NHS Trust in support of **Genomics England, East of England GMC**
  - led by Cambridge and including NUH Notts Trusts
  - Develop, integrate, host and support (at UHL) local instances of
    - CiviCRM patient cohort management
    - OpenClinica
    - Developing alongside Camb, Notts where possible
      - Phenotips for phenotyping (all sites)
      - Achiever for sample management (Notts+)

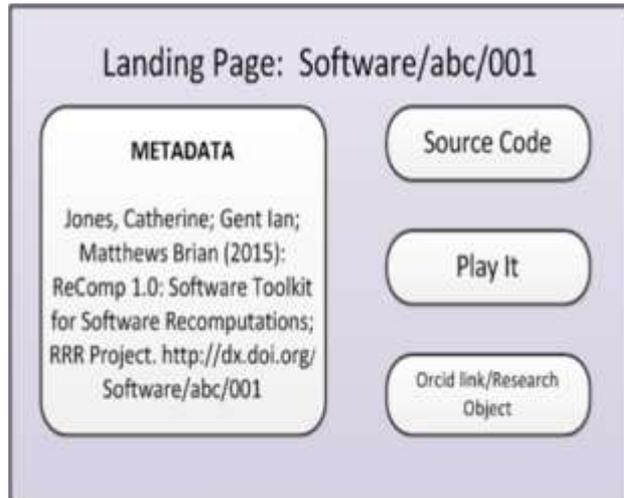
# BRISKit USPs

- *Integrated* support for core research processes
- Well-established mature open source applications as used in e.g. 100k Genomes NHS; Cardiovascular, Respiratory, Cancer: fully UK customised
- A cloud based platform for seamless management and integration between applications
- BRISKit RESTful APIs allow integration with existing clinical systems
- *Easy set up, use and administration through browser (including on mobile devices)*
- Capability of being *hosted in any compliant data centre provider*:
  - Aimes, Skyscape, UHL NHS Trust, SAIL (Farr) - NHS IGT compliant
  - MS Azure for Research - development, demo, anon

# Research Software Sustainability

- **Open Source community engagement**
    - standards compliance
    - consortium approach
    - work with grain of researchers
    - discipline specific forks?
    - Github versioning an example for research data?
  - UK Research Software Engineering is a long overdue developing resource and movement!
    - [www.rse.ac.uk](http://www.rse.ac.uk)
  - See Rob Baxter blog: “The research software engineer”
    - <http://dirkgorissen.com/2012/09/13/the-research-software-engineer/>
-

# Software RRR: Aims



Jisc funded **Software Re-use, Re-purposing and Reproducibility** project looked at how the DOI metadata schema should be applied to software written in an academic environment and explored capturing software in a running state.

Credit: Catherine Jones (STFC), Ian Gent (St Andrews), Jonathan Tedds (Leicester), Brian Matthews, Steven Lamerton, Tom Griffin and Paulina Lach (STFC)



Science & Technology  
Facilities Council

Log in | Register



**Tweets**

**mike galsworthy**  
@mikegalsworthy

#Opendata Hooray! New data journal just launched (Journal of Open Public Health Data) <http://t.co/ydmdINwPqM> @up\_jophd @openscience @UKODI

Jul 15th

**JOPHD**  
@up\_jophd

Just launched: the Journal of Open Public Health Data is now live! <http://t.co/JOL9uo4plyv>

Jul 15th

**Ubiquity Press** ]u[  
@ubiquitypress

New report out on realising an #OpenData culture from

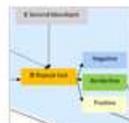
The Journal of Open Public Health Data (JOPHD) features peer reviewed data papers describing public health datasets with high reuse potential. We are working with a number of specialist and institutional data repositories to ensure that the associated data are professionally archived, preserved, and openly available. Equally importantly, the data and the papers are citable, and reuse is tracked.



### Most recent articles

 Alexander, Wint  
Projected Population Proximity Indices (30km) for 2005, 2030 & 2050  
11 Jul 2013

 Wint, Morley, Alexander  
Four Rodent and Vole Biodiversity Models for Europe  
11 Jul 2013

 Knowles, Olafsdottir  
Data: 'Initial Clinical Referral Standards after Newborn Screening...'  
09 Jul 2013

 Tadic, Knowles, Rahi  
Data from the Multiprofessional

### Most popular articles

 Alexander, Wint  
Projected Population Proximity Indices (30km) for 2005, 2030 & 2050  
11 Jul 2013

 Wint, Morley, Alexander  
Four Rodent and Vole Biodiversity Models for Europe  
11 Jul 2013

 Tadic, Knowles, Rahi  
Data from the Multiprofessional Workshop: 'Paediatric...'  
09 Jul 2013

 Knowles, Olafsdottir  
Data: 'Initial Clinical Referral Standards

Ads



**Open Access in developing countries**



**Latest:**

## **3. Aerosol Science for Public Health and Public Policy through Commercial Avenues**

**Dr. Josh Vande Hey**

NERC Knowledge Exchange Fellow

Air Quality Research Group / HARDi

University of Leicester

Email: [jvh7@le.ac.uk](mailto:jvh7@le.ac.uk)

Website:

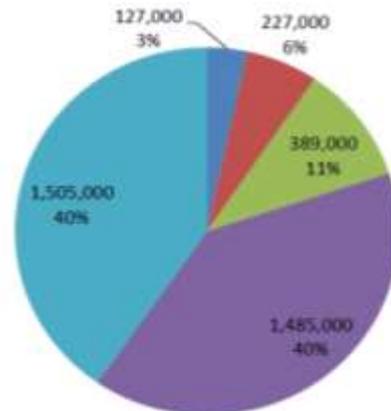
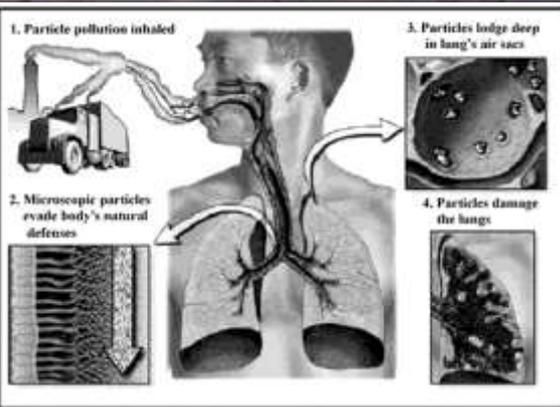
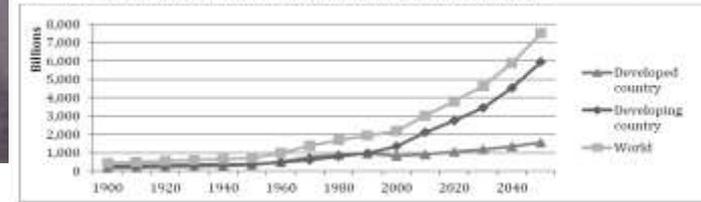
<https://www2.le.ac.uk/departments/physics/research/earth-observation-science/joshua-vande-hey>

# Why Aerosols?

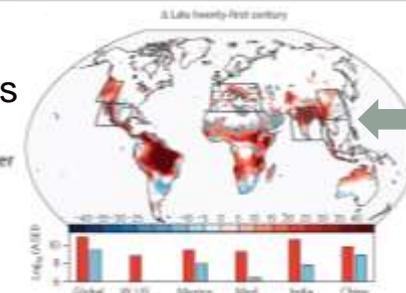
**£9-19 billion estimated annual economic cost of air pollution in the UK (HoCEAC).**



Global damage costs from air pollution, US Dollars (1990 prices) Hutton, 2011



WHO:  
AAP caused  
3.7 million deaths  
in 2012



Climate change  
and global  
development  
making air quality  
situation worse

AAP: Ambient air pollution; ALRI: Acute lower respiratory disease; COPD: Chronic obstructive pulmonary disease; IHD: Ischaemic heart disease.

DEFRA: The particulate pollution burden in the UK was estimated to be equivalent to **29,000 deaths** and **340,000 life years lost** in **2008 alone**.

China is investing **>£90 billion** to reduce PM2.5 levels in Beijing by 25% **by 2017**.



Healthy lung      Lung suffering from Pneumoconiosis



\*Circled partners linked to letters of support

KNOWLEDGE BASE:  
AEROSOL SOURCES,  
CHEMISTRY, TRANSPORT  
AND MEASUREMENT

1

STAKEHOLDER NEED:  
MANAGING HEALTH  
IMPACT AROUND  
PERSONAL EXPOSURE



KNOWLEDGE BASE:  
AEROSOL SOURCES,  
CHEMISTRY, TRANSPORT  
AND MEASUREMENT

1

STAKEHOLDER NEED:  
MANAGING HEALTH  
IMPACT AROUND  
PERSONAL EXPOSURE

2

MARKET AND SYSTEMS  
FEASIBILITY: INTEGRATION  
OF NEW TECHNOLOGY AND  
DATA STREAMS



# Some thoughts

- Can't do it all in house but of course need specialist work!
- But many disciplines don't have data centres
- Build coalition of institutional actors
  - Essential to have high level support
- Take and shape
  - Identify what you do have in-house
  - Access external tools, standards where possible
    - Active storage, collaboration, eprints...
  - Propose best of breed for (inter)national reuse
    - Share benefits (and costs) over academic networks
- Sustainability the key challenge
- As much cultural as technical - needs networks...
- Make use of e.g. DCC expertise and resources!

# It is worth it & funders are paying attention!



- increase the trustworthiness and value of individual data sets
- strengthen the findings based on cited data sets
- increase the transparency and traceability of data and publications
- enable reuse and repurposing



Thank you for listening and thanks to UoL  
and project teams, partners 😊

Dr Jonathan Tedds [jat26@le.ac.uk](mailto:jat26@le.ac.uk) @jtedds

Senior Research Fellow,  
Dept of Health Sciences  
University of Leicester

#BRISKit <http://www.brisskit.le.ac.uk>

Director: RED Informatics Ltd @REDinfx

**What is BRISKit?**

**SEE 2 MINUTE ANIMATED EXPLAINER !**

<https://www.brisshit.le.ac.uk/node/61>