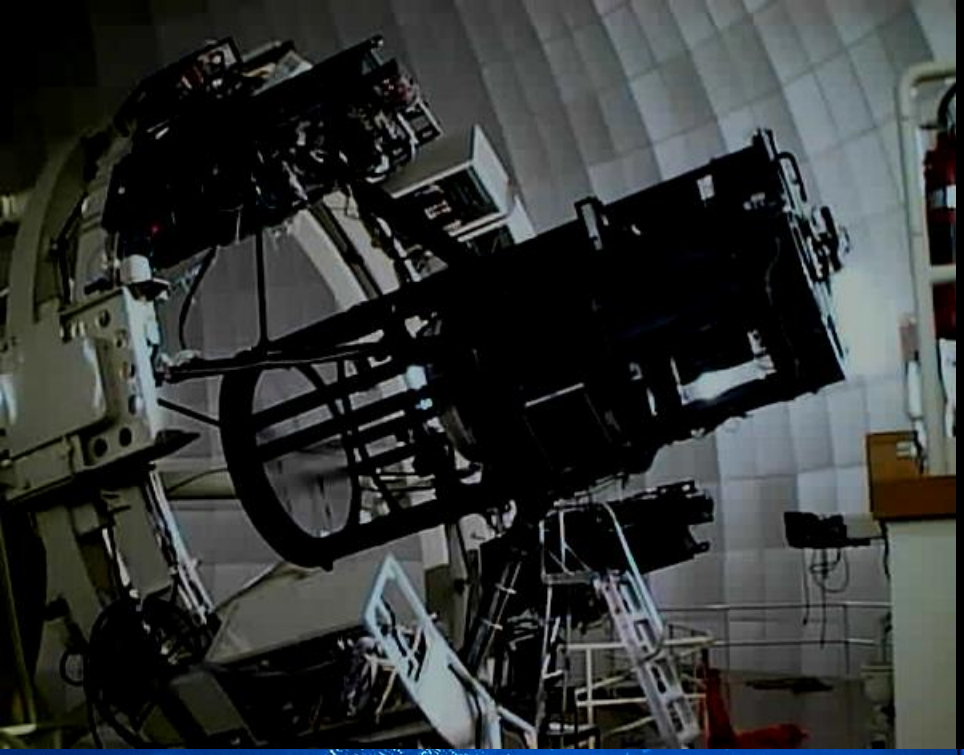
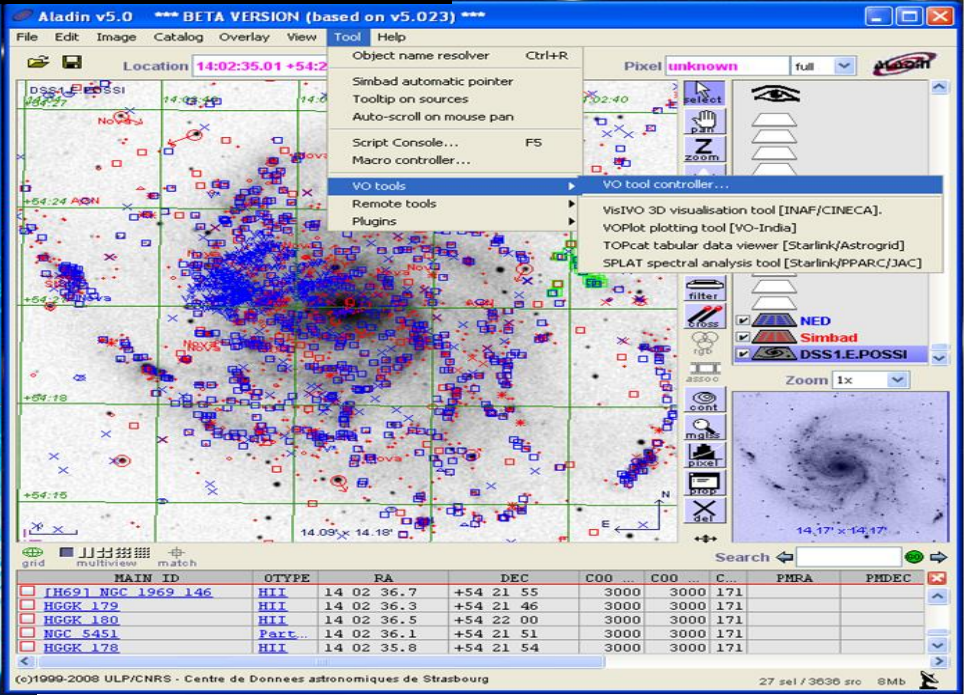
Things to do with data: Biomedical research databases & data sharing

Dr Jonathan Tedds **@jtedds**
PI BRISSKit
Senior Research Fellow
Health & Research Data Informatics
University of Leicester

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

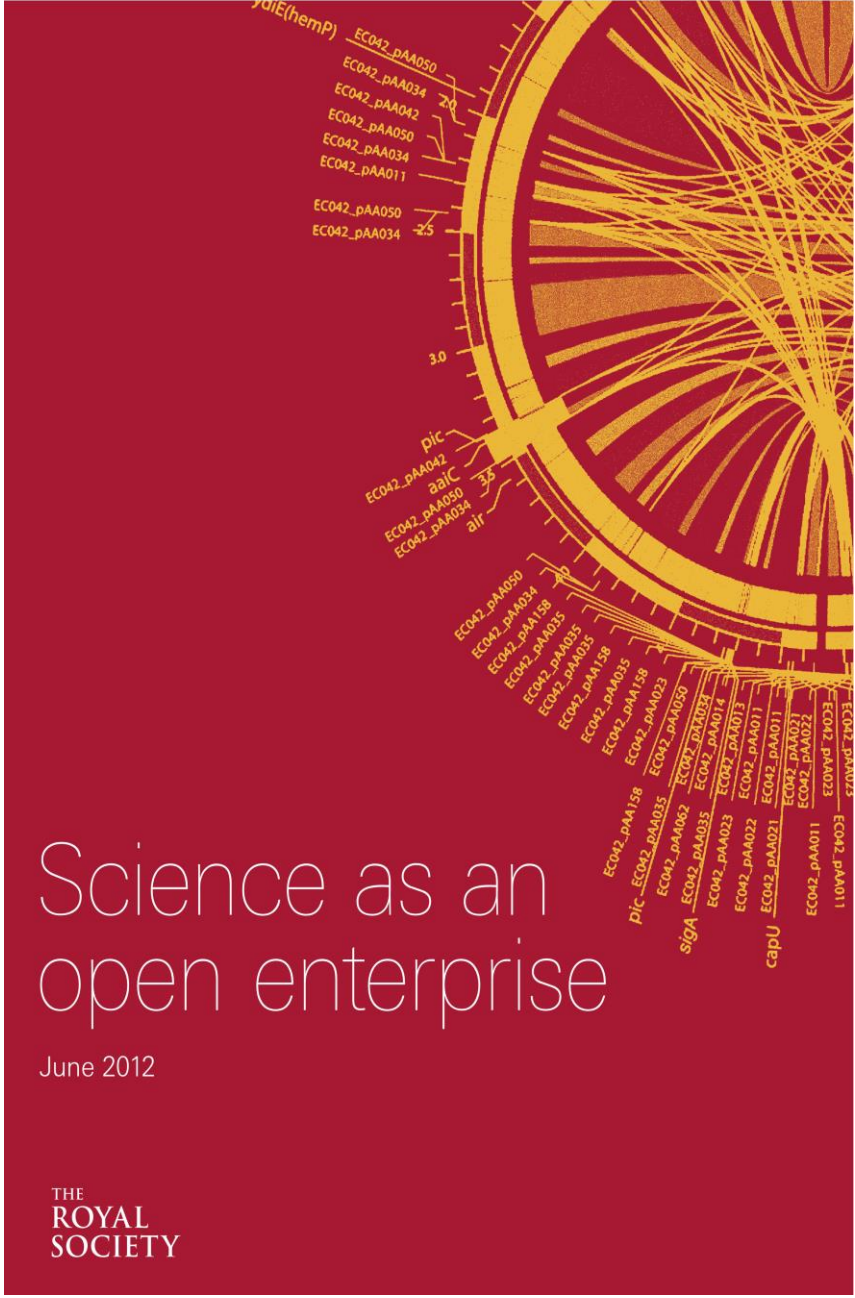
Oxford, 3rd Dec 2014



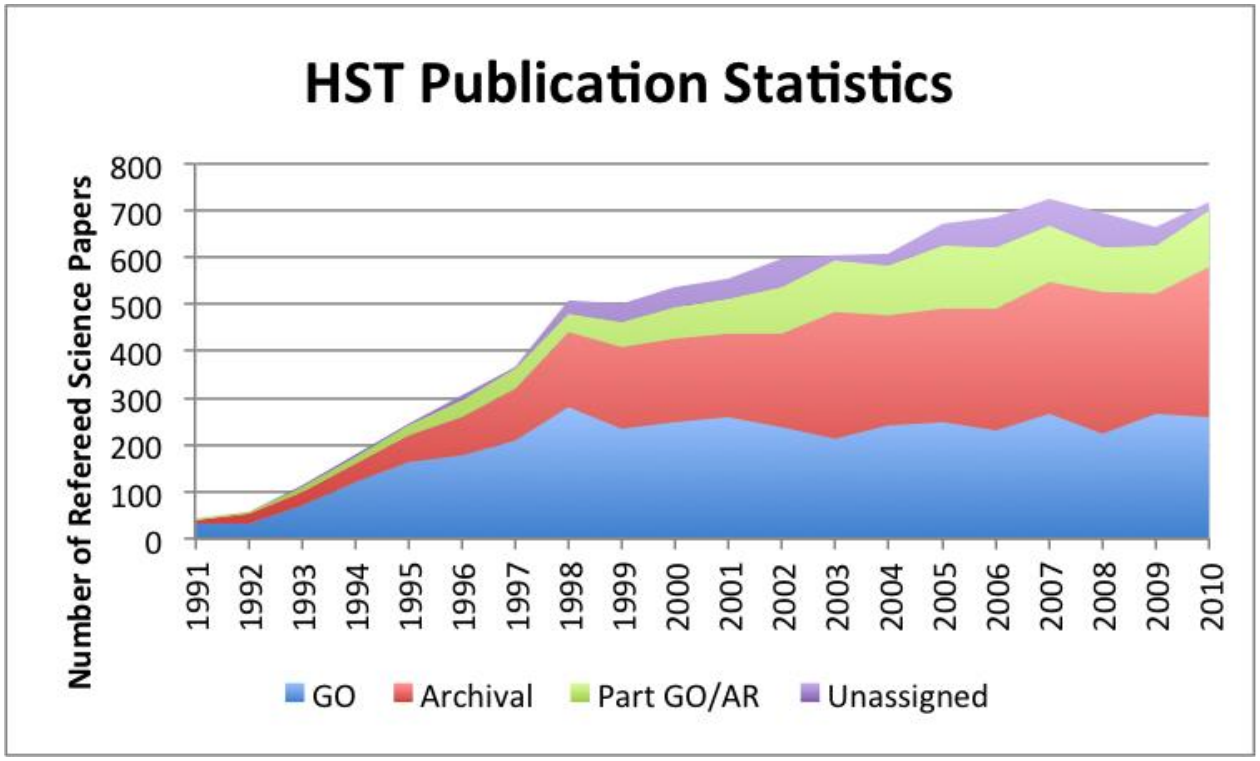

Aladin v5.0 *** BETA VERSION (based on v5.023) ***
File Edit Image Catalog Overlay View Tool Help
Object name resolver Ctrl+R
Simbad automatic pointer
Tooltip on sources
Auto-scroll on mouse pan
Script Console... F5
Macro controller...
VO tools
Remote tools
Plugins
VO tool controller...
VisIVO 3D visualisation tool [INAF/CINECA].
VOPlot plotting tool [VO-India]
TOPcat tabular data viewer [Starlink/Astrogrid]
SPLAT spectral analysis tool [Starlink/PPARC/JAC]
Location 14:02:35.01 +54:2
Pixel unknown full
NED
Simbad
DSS1.E.POSSI
Zoom 1x
Search
MAIN ID OTYPE RA DEC C00 ... C00 ... C... PMRA PMDEC
[H69] NGC 1969 146 HII 14 02 36.7 +54 21 55 3000 3000 171
HGGK 179 HII 14 02 36.3 +54 21 46 3000 3000 171
HGGK 180 HII 14 02 36.5 +54 22 00 3000 3000 171
NGC 5451 Part... 14 02 35.1 +54 21 51 3000 3000 171
HGGK 178 HII 14 02 35.8 +54 21 54 3000 3000 171
(c)1999-2008 ULP/CNRS - Centre de Donnees astronomiques de Strasbourg
27 sel / 3636 src 8Mb

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/

Oh, and…. G8 UK says so :P

UNITED KINGDOM 2013

We are committed to openness in scientific research data to speed up the progress of scientific discovery, create innovation, ensure that the results of scientific research are as widely available as practical, enable transparency in science and engage the public in the scientific process.

- To the greatest extent and with the fewest constraints possible publicly funded scientific research data should be open, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests, whilst acknowledging the legitimate concerns of private partners.

- Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards.

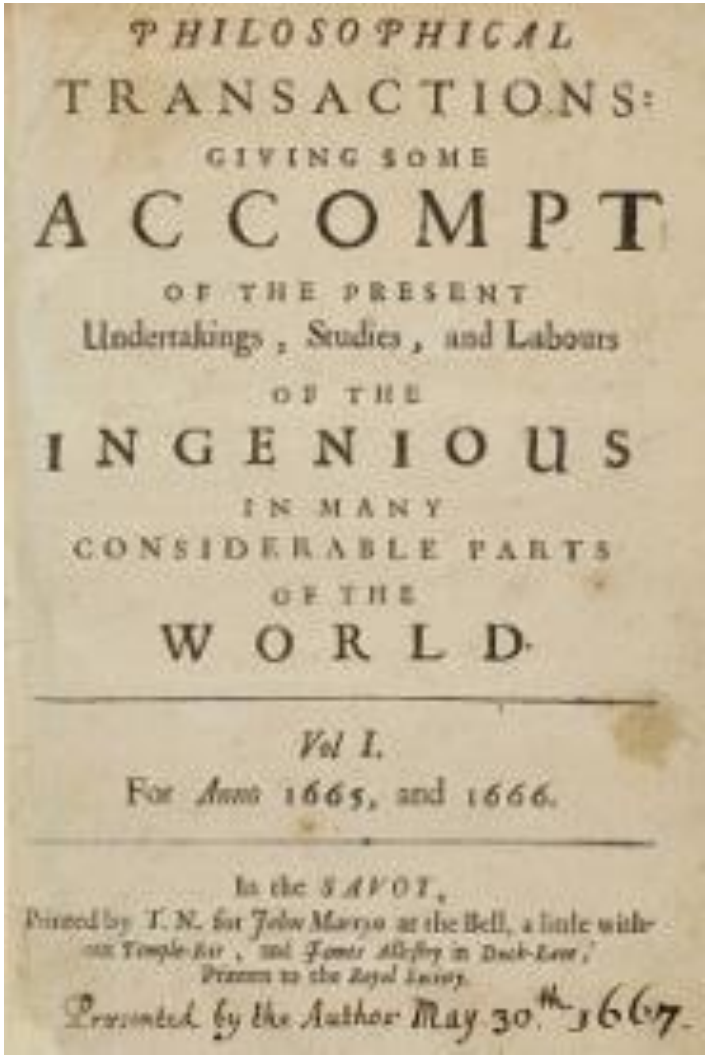
- To ensure successful adoption by scientific communities, open scientific research data principles will need to be underpinned by an appropriate policy environment, including recognition of researchers fulfilling these principles, and appropriate digital infrastructure.

Scale of the problem: who, what, when where….?

http://blogs.scientificamerican.com/absolutely-maybe/2013/09/10/opening-a-can-of-data-sharing-worms/

- Timothy Vines and colleagues studied reproducibility of data sets in zoology and changes through time
  - gathered 516 papers published between 1991 and 2011
  - then they tried to track the data down…
- Even tracking down the authors was a challenge
  - Over time a dwindling minority of papers were accompanied by author email addresses that still functioned
- only 37% of the data - even from papers in 2011 - were still findable and retrievable
  - proportion dropped each earlier year
- For papers published in 1991
  - only 7% of the data could be determined to truly still be in existence and retrievable
  - few authors could be found, and most of them were reporting that their data were lost or inaccessible

This isn't new…


PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.

Vol I.
For Anno 1665, and 1666.

In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little within
the Temple-Bar, and James Allestry in Duck-Lane,
Printers to the Royal Society.
Presented by the Author May 30th 1667.

Henry Oldenburg

- inveterate correspondent
- now think of as scientist

- Had idea to publish Philosophical Transactions (1665):
  - Should be written in vernacular not Latin
  - Underlying evidence must be concurrently published
  - Helped propel Europe at the time
  - Concept of scientific self correction
    - able to write it's errors
- Wrote: "thought fit to employ the [printing] press……..Universal Good of Mankind"
  - How do we achieve these ends in the post-Gutenburg era?

Data as a "public good" (2011)

RESEARCH COUNCILS UK

Excellence with Impact

Accessibility

Research and funding

Research Careers

Public Engagement with Research

Knowledge Exchange and Impact

International

Press and Media

Publications

About

Home

Research and Funding

Research Funding

Areas of Research

Cross-Council Research Themes

Research Infrastructure

Research Priorities

Peer review

Eligibility for Research Council funding

How to apply for research funding

Applications which may cross Research Council remits

Terms and Conditions of Research Council fEC Grants

Terms and Conditions of Research Council Training Grants

Open Access

RCUK Common Principles on Data Policy

Efficiency

Home > Research and Funding > RCUK Common Principles on Data Policy

RCUK Common Principles on Data Policy

Making research data available to users is a core part of the Research Councils' remit and is undertaken in a variety of ways. We are committed to transparency and to a coherent approach across the research base. These RCUK common principles on data policy provide an overarching framework for individual Research Council policies on data policy.

Principles

- Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.
- Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice. Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.
- To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.
- RCUK recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.
- To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The length of this period varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils.
- In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.
- It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.

Search:

Go

This website

All Research Councils

- Public good
- Preservation
- Discovery
- Confidentiality
- First use
- Recognition
- Public funding

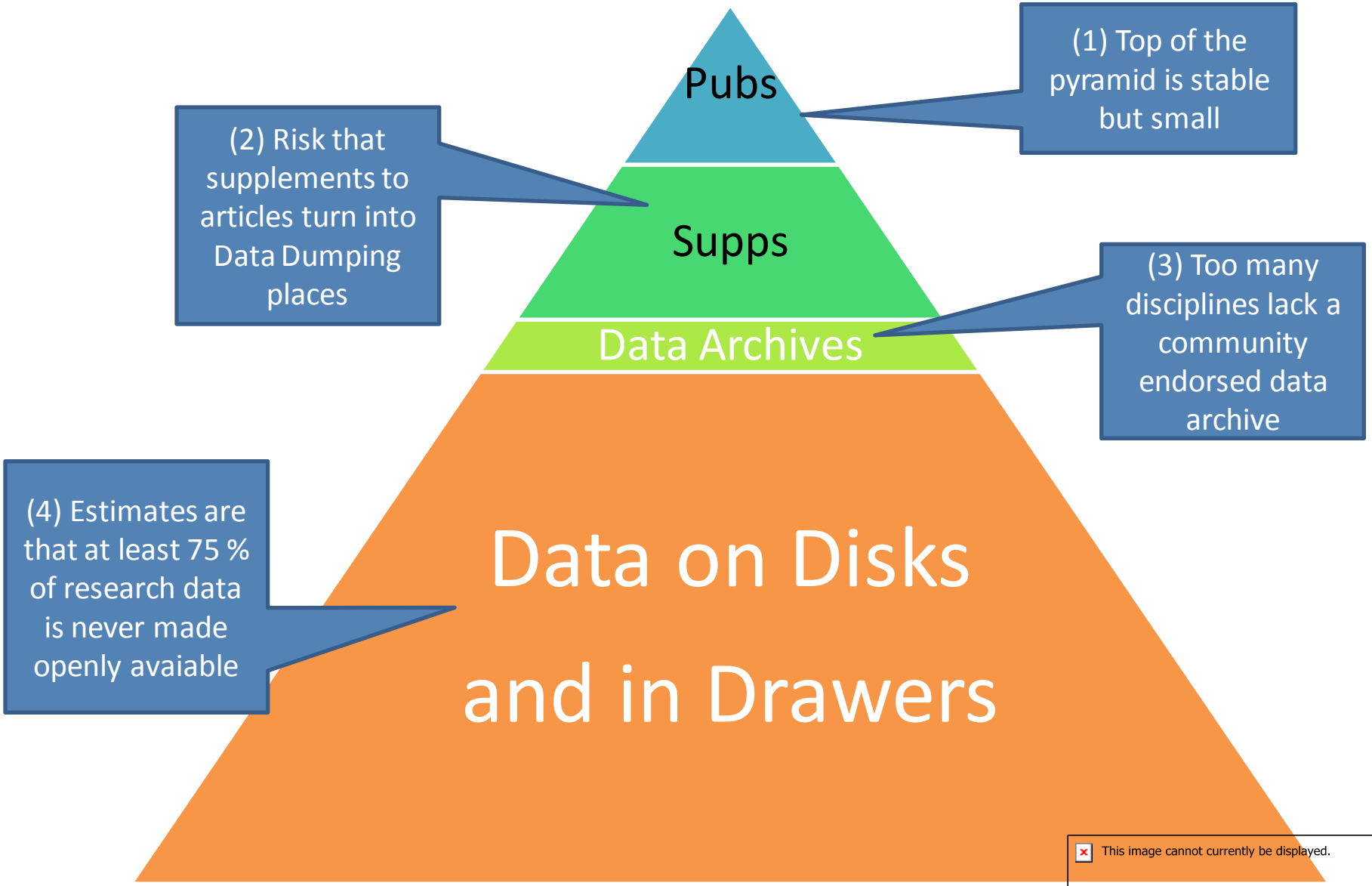
So what do we mean by publishing data?

- The familiar:
  - Supplementary tables via journal or
  - Archived raw or calibrated facility data
  - Discipline specific and institutional / national archives

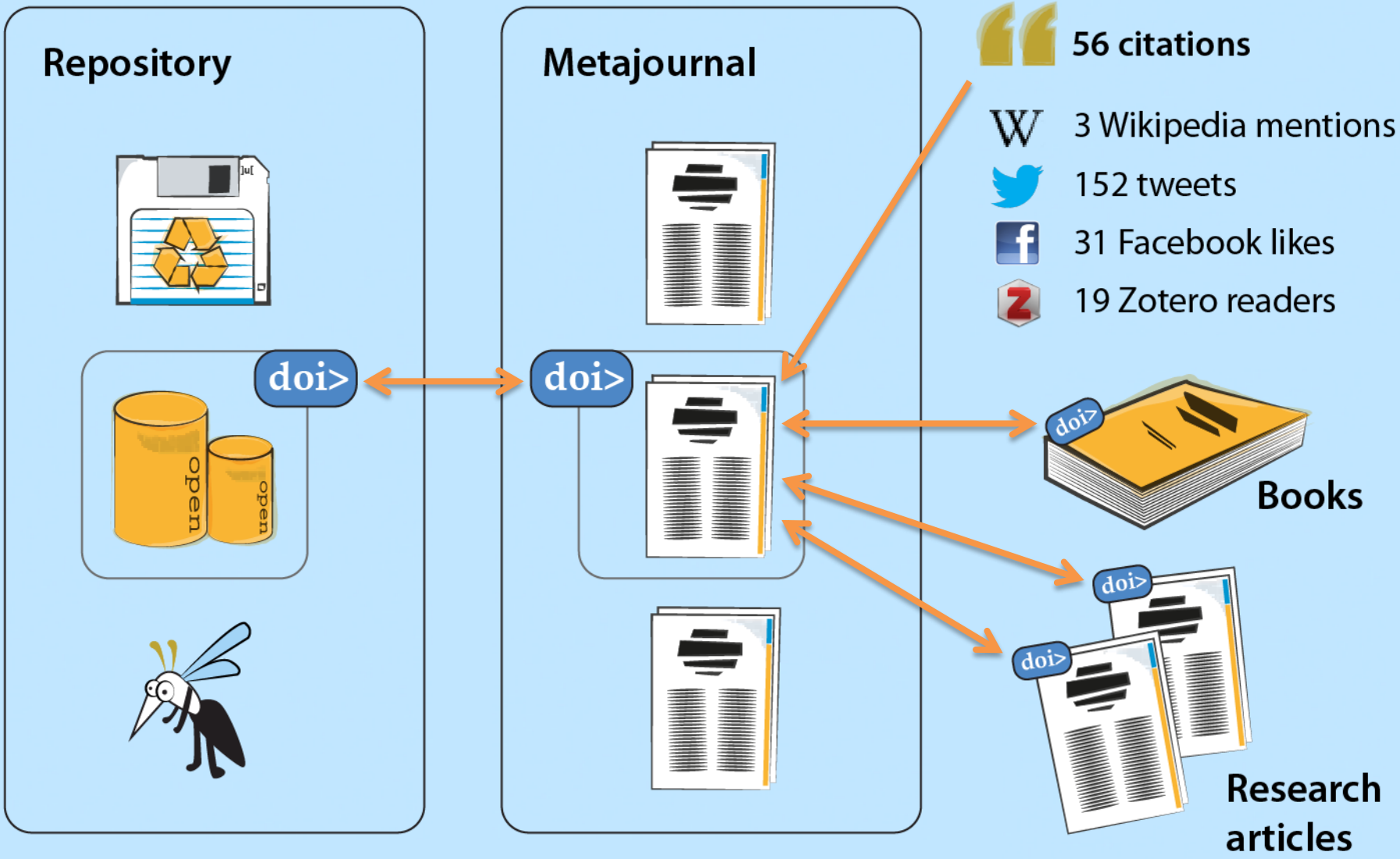
- Data under the graph?
  - In order to reproduce and adapt article analysis

- “Research ready” open data
  - In order to reuse and repurpose
  - for interdisciplinary researchers, community, business
  - Ideally peer reviewed?

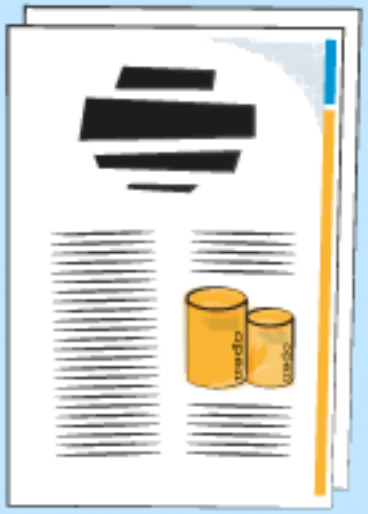
ODE Data Publication Pyramid:



This image cannot currently be displayed.


Repository
Metajournal
56 citations
3 Wikipedia mentions
152 tweets
31 Facebook likes
19 Zotero readers
doi>
doi>
open
open
doi>
Books
doi>
doi>
Research articles

Structure of a data paper:



A data paper…

- … describes the methodology with which a dataset was created.
- … describes the dataset itself.
- … details the reuse potential of the data.
- … is often authored by a data scientist.
- … is citable, enabling reuse to be tracked.

A data paper is <u>not…</u>

- … a research paper. A data paper only describes a dataset. But it will reference research papers that are based on the data.
- … simply replication of the information in a data repository.

Peer review

1. The paper contents

a. The methods section of the paper must provide sufficient detail that a reader can understand how the resource was created.
b. The resource must be correctly described.
c. The reuse section must provide concrete and useful suggestions for reuse of the reuse.

2. The deposited resource

a. The repository must be suitable for resource and have a sustainability model.
b. Open license permits unrestricted access if possible (e.g. CC0), or guarantees controlled access if unavoidable.
c. A version in an open, non-proprietary format.
d. Labeled in such a way that a 3rd party can make sense of it.
e. Must be actionable.

UPmetajournals

open health data

]u[ubiquity press open access

Home | About | Editorial Board | Contact | Repositories | FAQ | Articles | Archive

Log in | Register

Follow on Twitter

Follow via RSS

A A A

Tweets

mike galsworthy
@mikegalsworthy

#Opendata Hooray! New data journal just launched (Journal of Open Public Health Data) http://t.co/ydmdlNwPqM @up_jophd @openscience @UKODI

Jul 15th

JOPHD
@up_jophd

Just launched: the Journal of Open Public Health Data is now live! http://t.co/OL9uo4plyv

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

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
- *Welcoming and reasonable journal homes!*

PREPARDE: Peer REview for Publication & Accreditation of Research Data in the Earth sciences

Jonathan Tedds (Leicester), Sarah Callaghan (BADC), Fiona Murphy (Wiley), Rebecca Lawrence (F1000R), Geraldine Stoneham (MRC), Elizabeth Newbold (BL), Rachel Kotarski (BL), Matthew Mayernik (NCAR), John Kunze, Carly Strasser (CDL), Angus Whyte (DCC), Becca Wilson (Leicester), Simon Hodson (Jisc) and #PREPARDE project team

+ Geraldine Clement Stoneham (MRC), Elizabeth Newbold, Rachel Kotarski (BL) on data peer review

http://www.le.ac.uk/projects/preparde

From Mayernik et al. 2014 Most cited Bulletin of the American Meteorological Society (BAMS) articles. Data from Web of Science, gathered on June 11, 2013

http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-13-00083.1

Article	Data paper?	Citations	Article details
1	Yes	10,113	Kalnay, E; et al. The NCEP/NCAR 40-year reanalysis project, 1996.
2	No	3,201	Torrence, C; Compo, GP. A practical guide to wavelet analysis, 1998.
3	No	2,367	Mantua, NJ; et al. A Pacific interdecadal climate oscillation with impacts on salmon production, 1997.
4	Yes	1,987	Kistler, R; et al. The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, 2001.
5	Yes	1,791	Xie, PP; Arkin, PA. Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, 1997.
6	Yes	1,448	Kanamitsu, M; et al. NCEP-DOE AMIP-II reanalysis (R-2), 2002.
7	No	1,014	Baldocchi, D; et al. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, 2001.
8	Yes	902	Rossow, WB; Schiffer, RA. Advances in understanding clouds from ISCCP, 1999.
9	Yes	900	Rossow, WB; Schiffer, RA. ISCCP cloud data products, 1991.
10	No	877	Hess, M; Koepke, P; Schult, I. Optical properties of aerosols and clouds: The software package OPAC, 1998.
11	No	815	Willmott, CJ. Some comments on the evaluation of model performance, 1982.
12	No	815	Trenberth, KE. The definition of El Nino, 1997.
13	Yes	785	Woodruff, SD; Slutz, RJ; et al. A comprehensive ocean-atmosphere data set, 1987.
14	Yes	776	Meehl, G.A.; et al. The WCRP CMIP3 multimodel dataset - A new era in climate change research, 2007.

Enabling Open Data Publishing

- 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

HALOGEN (History, Archaeology, Linguistics, Onomastics, GENetics):

Throwing light on the past through cross-disciplinary databasing

University of Leicester

wellcome trust

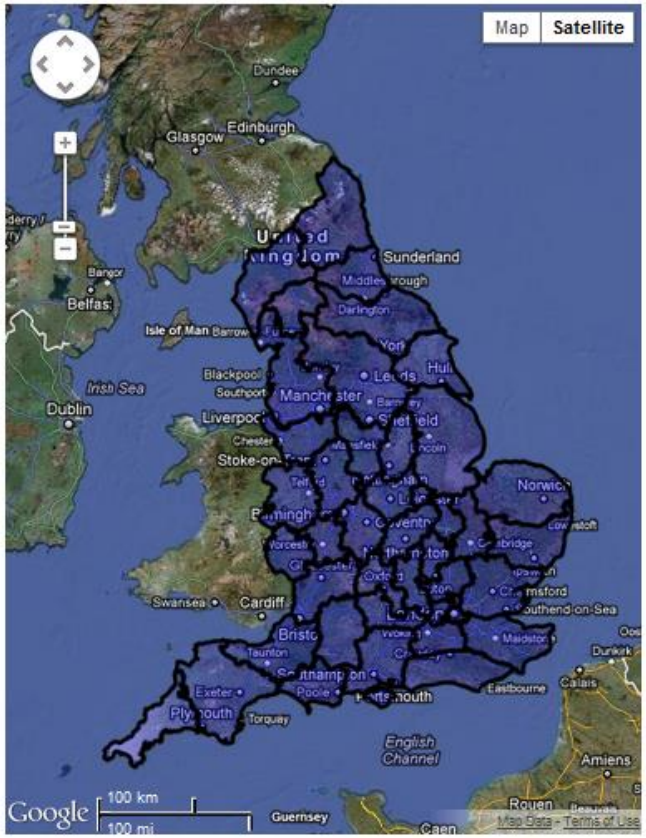
- Portable Antiquities Scheme (British Museum)
- Place-names (Nottingham)
- Surnames
- Genetics
- IT hosting and GIS
- Best practice: #JISCMRD, UKRDS, DCC, international

http://halogen.le.ac.uk

JISC

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
- -dwostle
- -ek
- -ell

Search by place name:

Exact match

Submit Reset

University of Leicester

JISC The University of Nottingham THE BRITISH MUSEUM

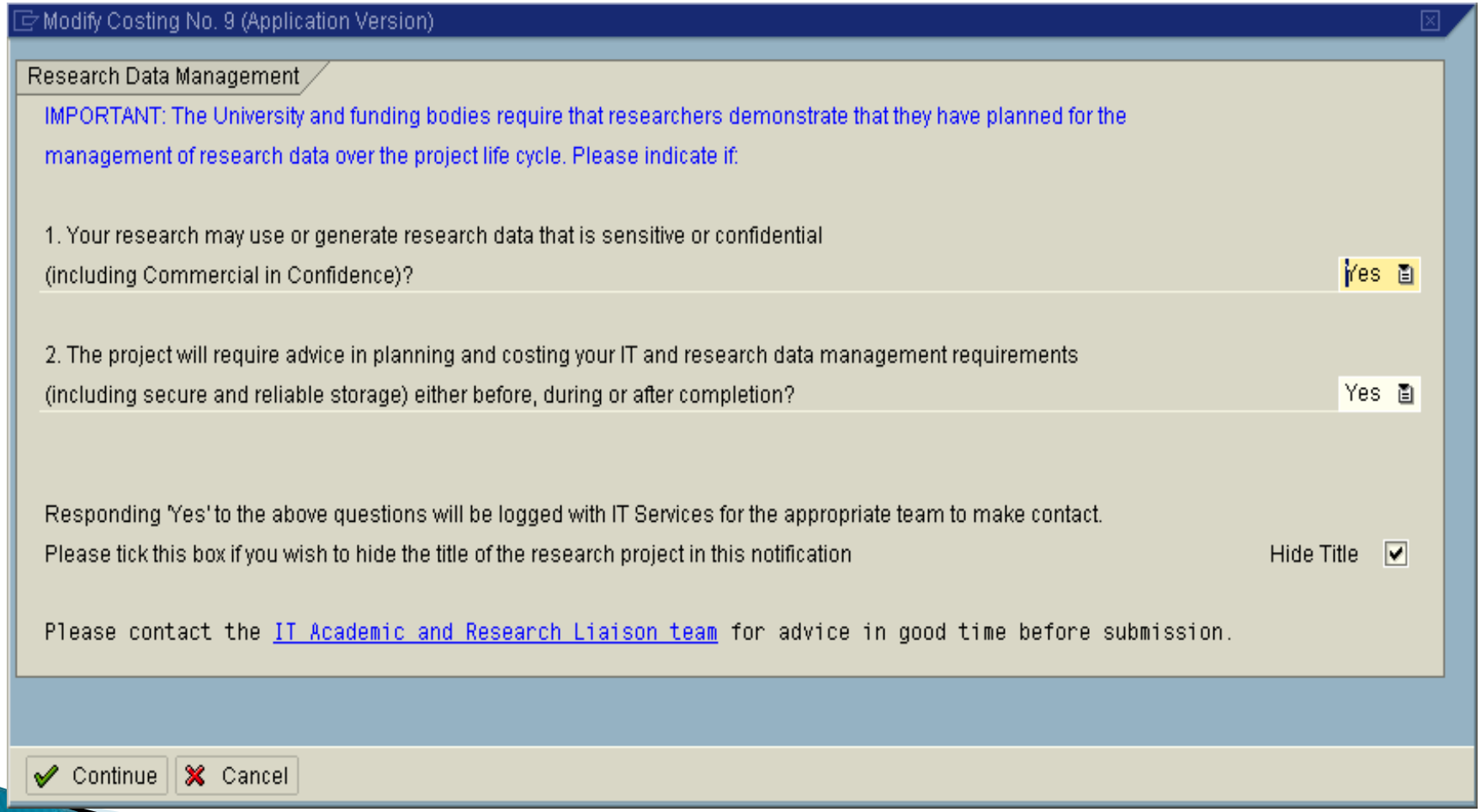
Benefits

- *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*

University *of* Leicester

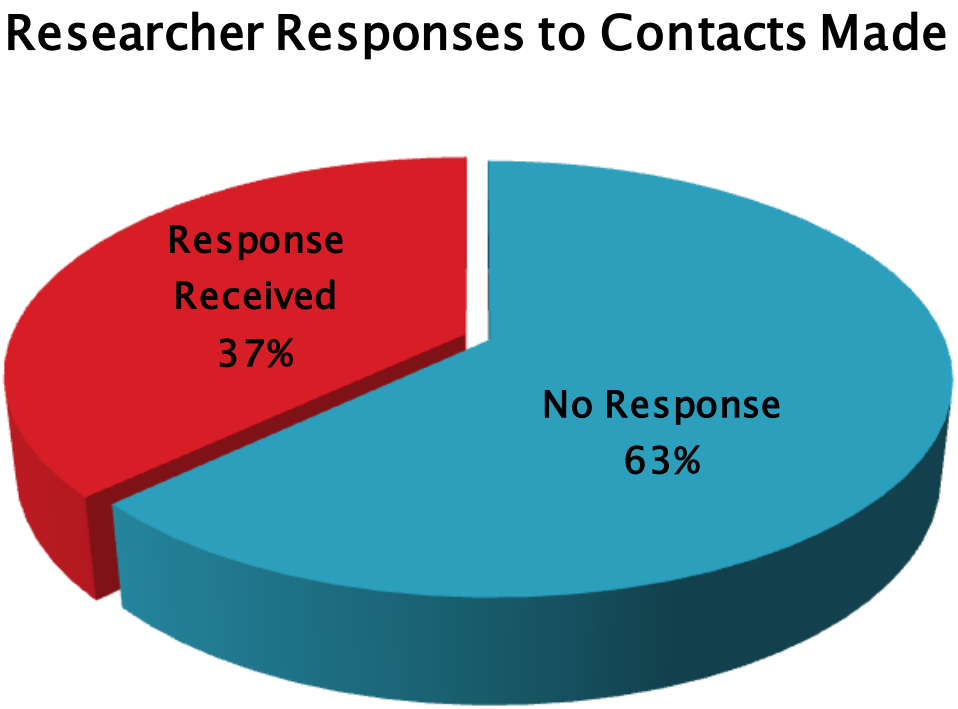
Reward = Leverhulme Trust funding £1.3m!

Top Tip: how to get researchers' attention?



University of Leicester

Research costing – only part of the answer



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?**

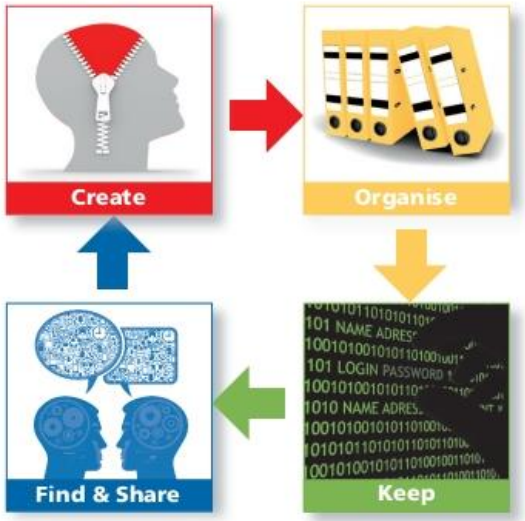
  Is there an appropriate disciplinary or institutional repository and what do you need to do to deposit your research output? Plan ahead to avoid refactoring...

- [ ] **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**



email: researchdata@le.ac.uk

University of Leicester

University of Leicester

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.*

www.le.ac.uk/researchdata

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***?

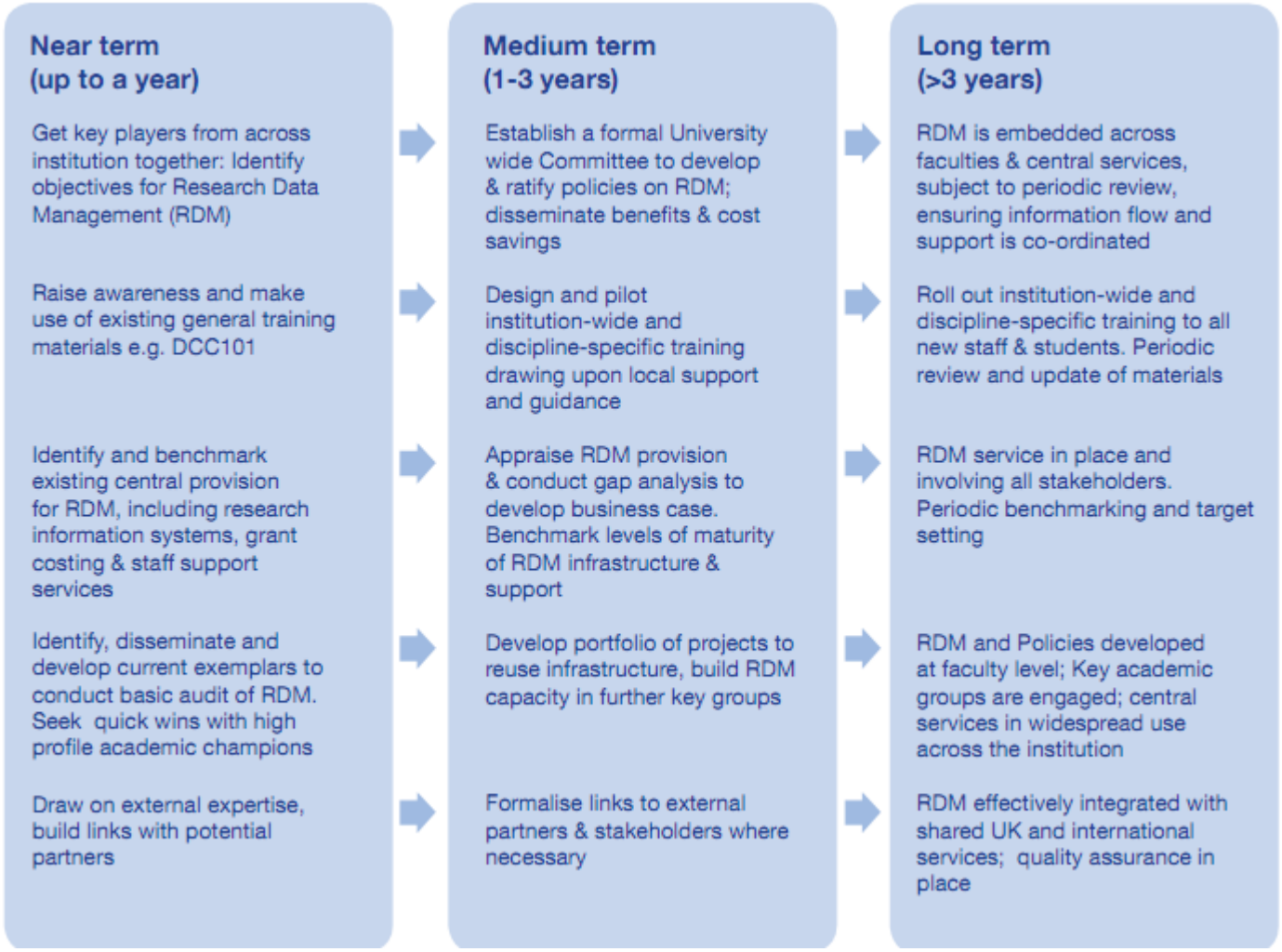
- [ ] **how securely your data is stored?**

  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



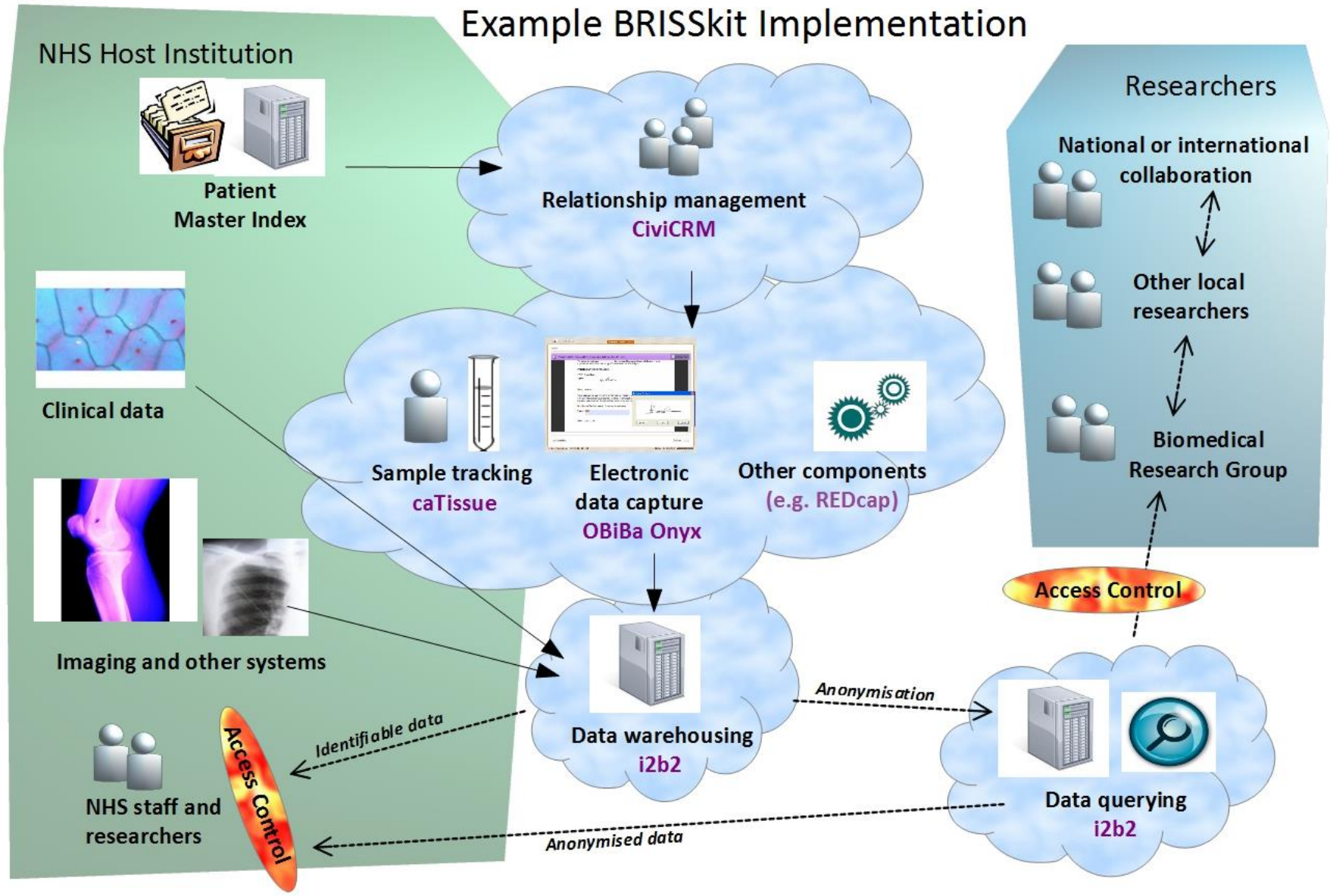
From Whyte & Tedds (2011), DCC Briefing
http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm

Challenge for institutions

- Rise to scientific and research challenge
  - Not just a management challenge
  - Responsibility for the knowledge they create
- Library
  - “Doing the wrong things through the wrong people”?
  - Challenge for library to enable:
    - 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

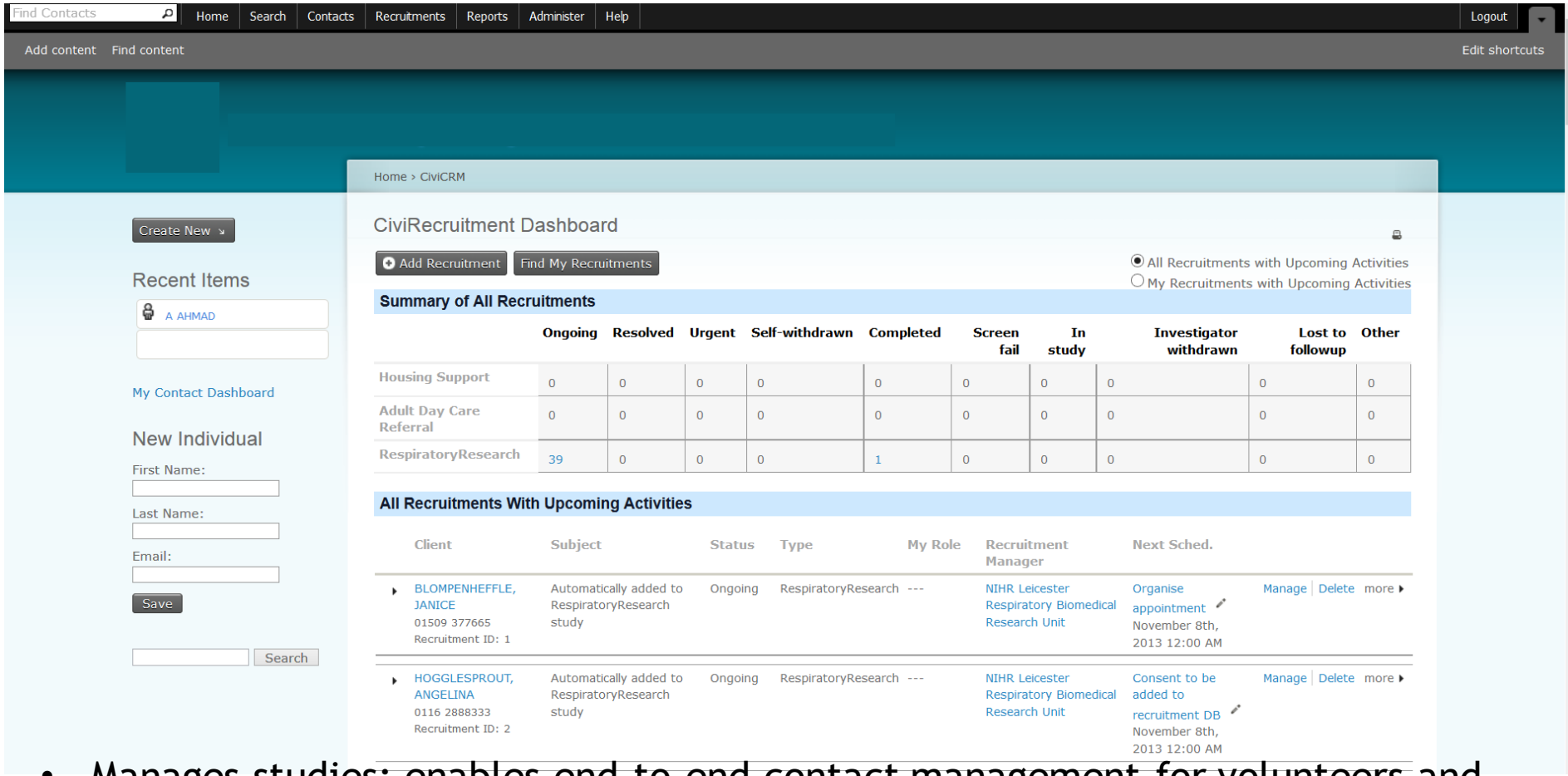
Enabling Open Data Publishing

- 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

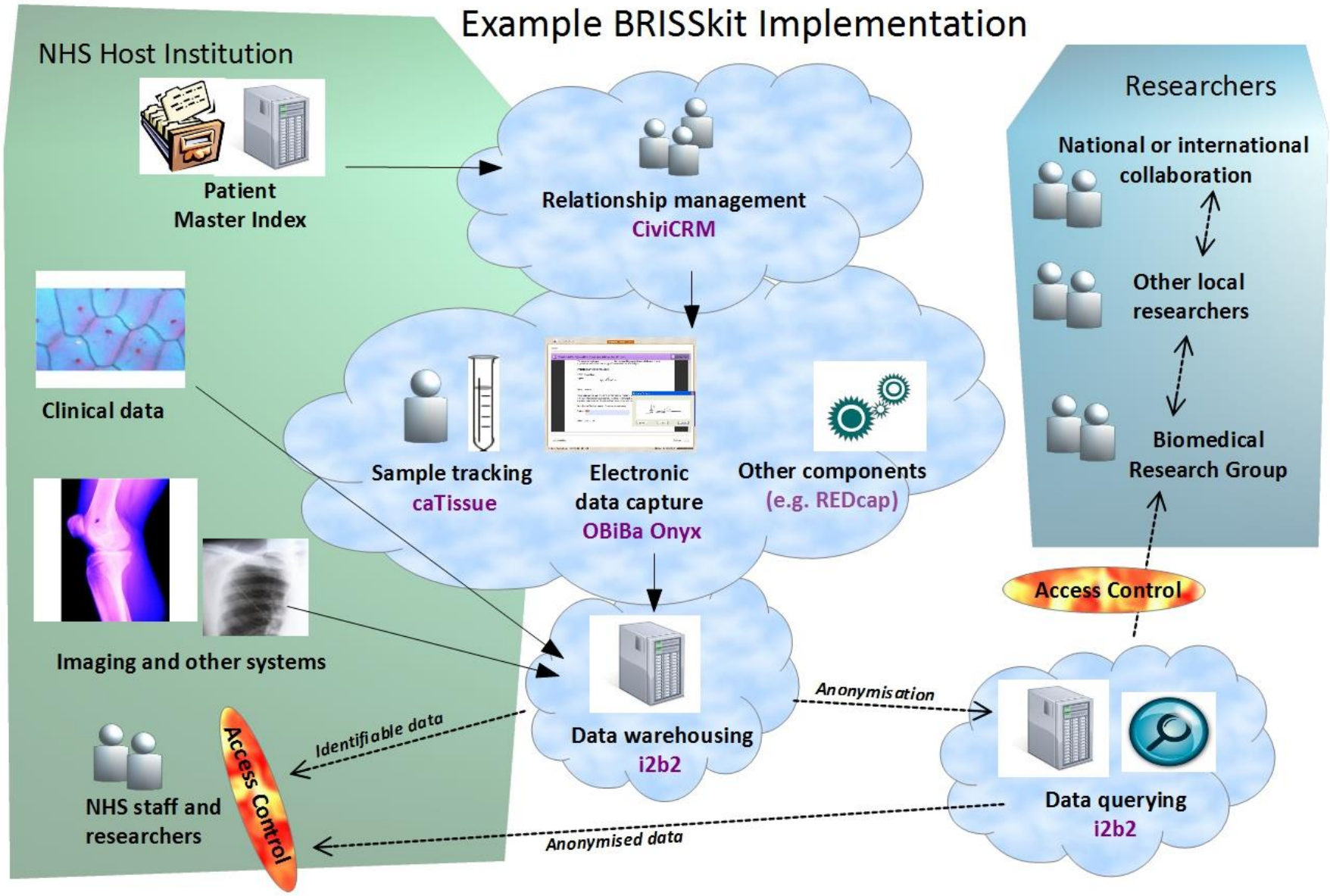


http://www.brisskit.le.ac.uk

BRISSKit CiviCRM: patient cohort management

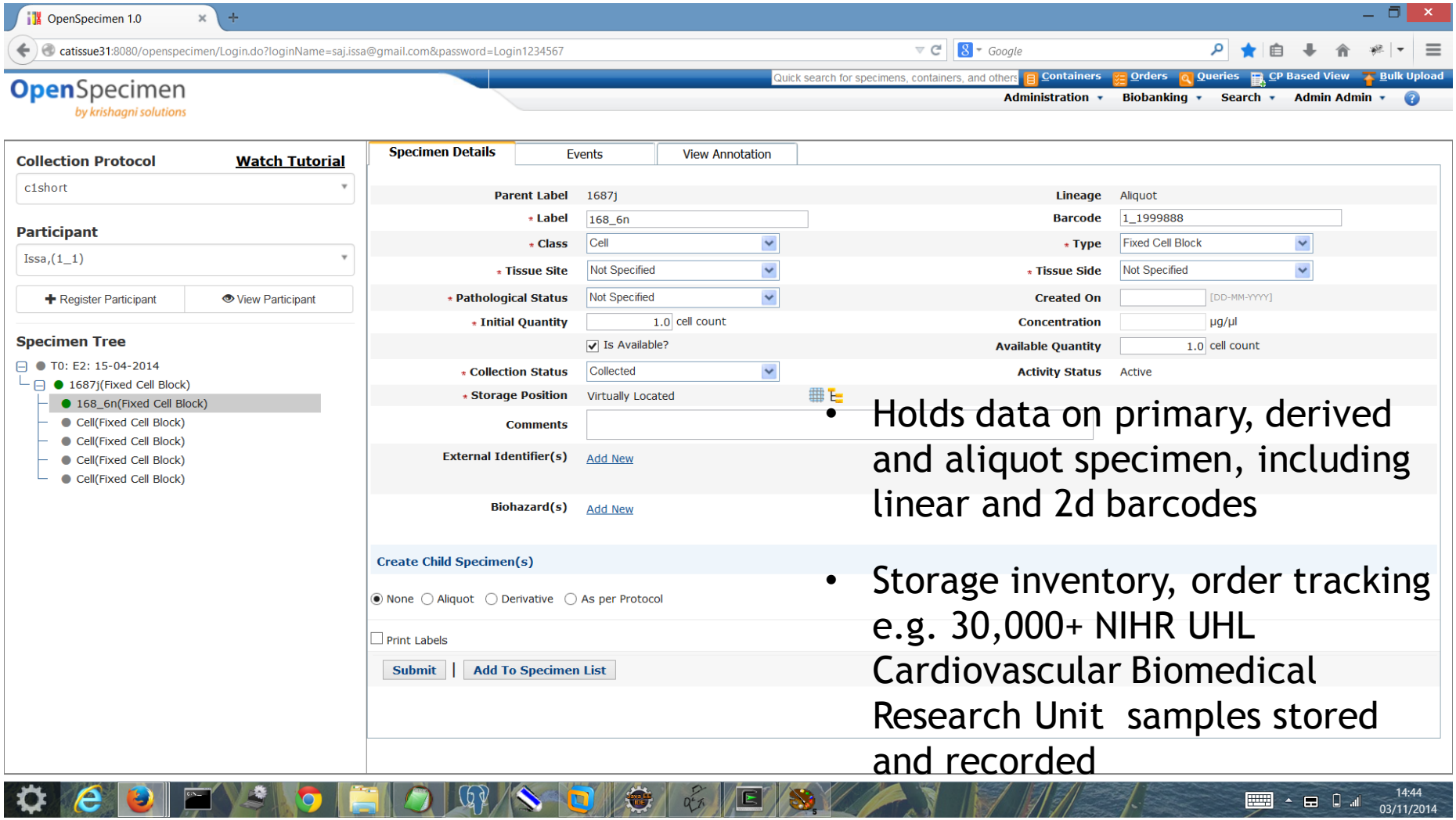


- Manages studies: enables end-to-end contact management for volunteers and research participants
- track approaches, contact, responses, recruitment, exclusions
- object model that reflects community building and non-profit relationships



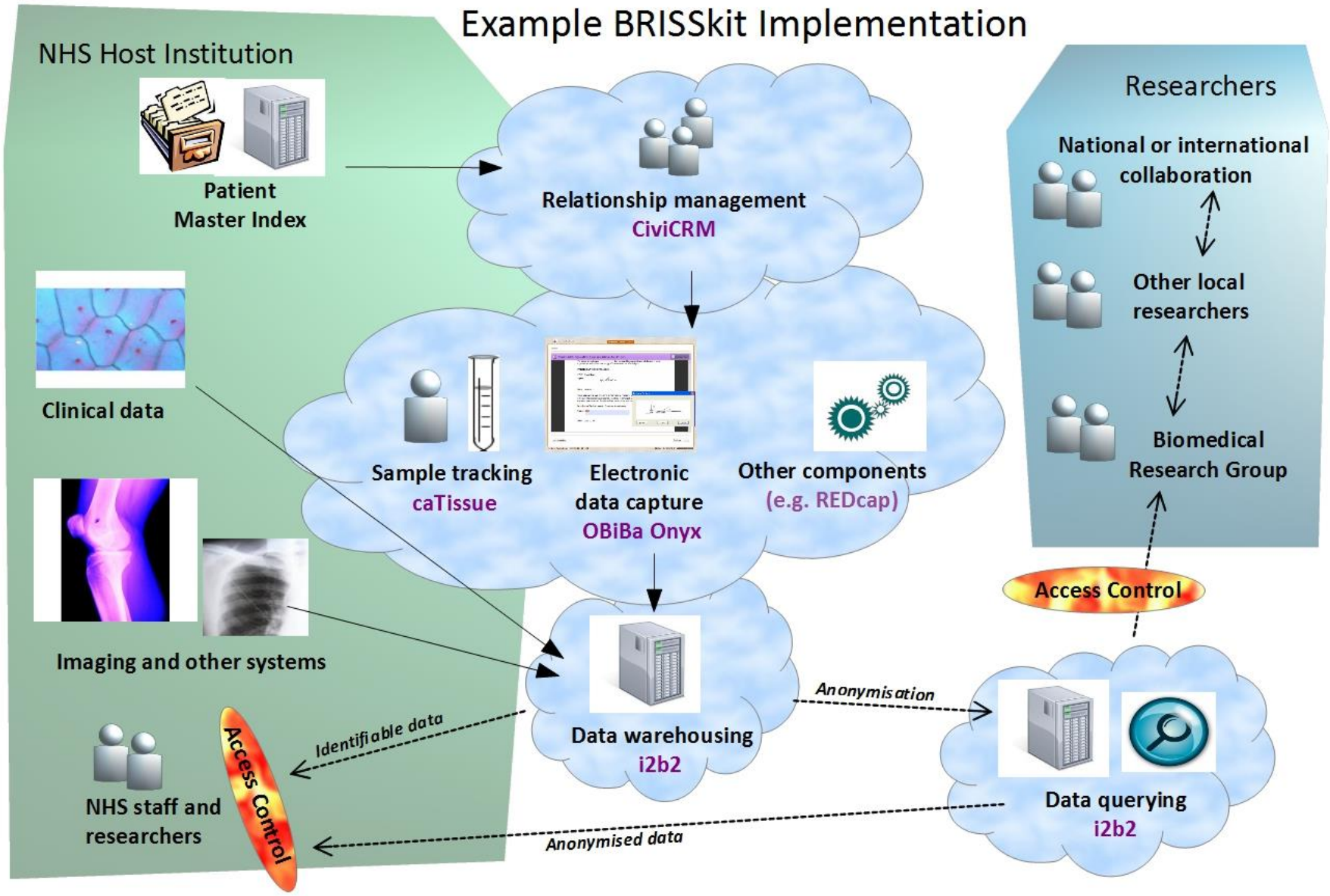
http://www.brisskit.le.ac.uk

BRISSKit OpenSpecimen: sample management



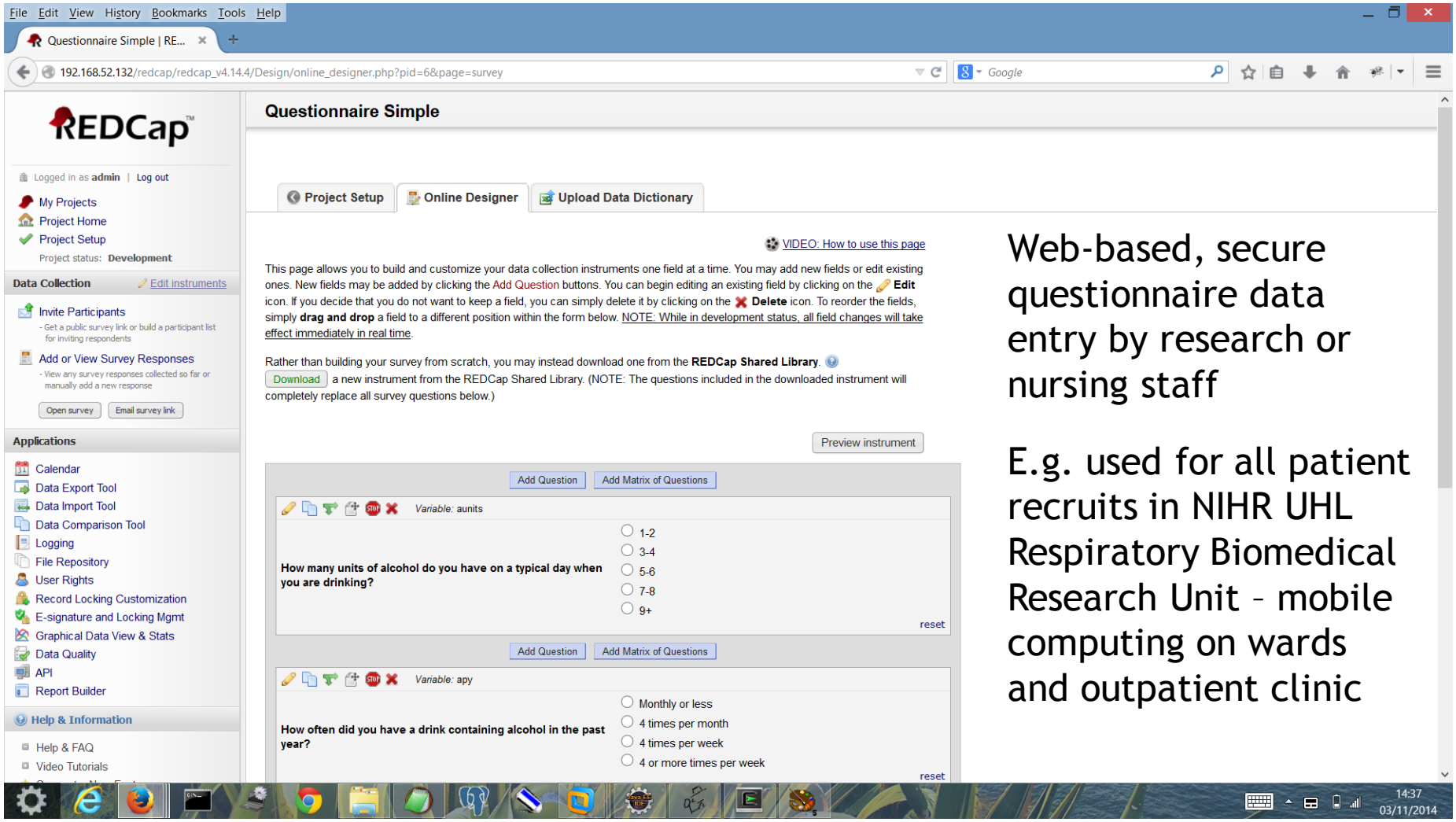
- Holds data on primary, derived and aliquot specimen, including linear and 2d barcodes
- Storage inventory, order tracking e.g. 30,000+ NIHR UHL Cardiovascular Biomedical Research Unit samples stored and recorded

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



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

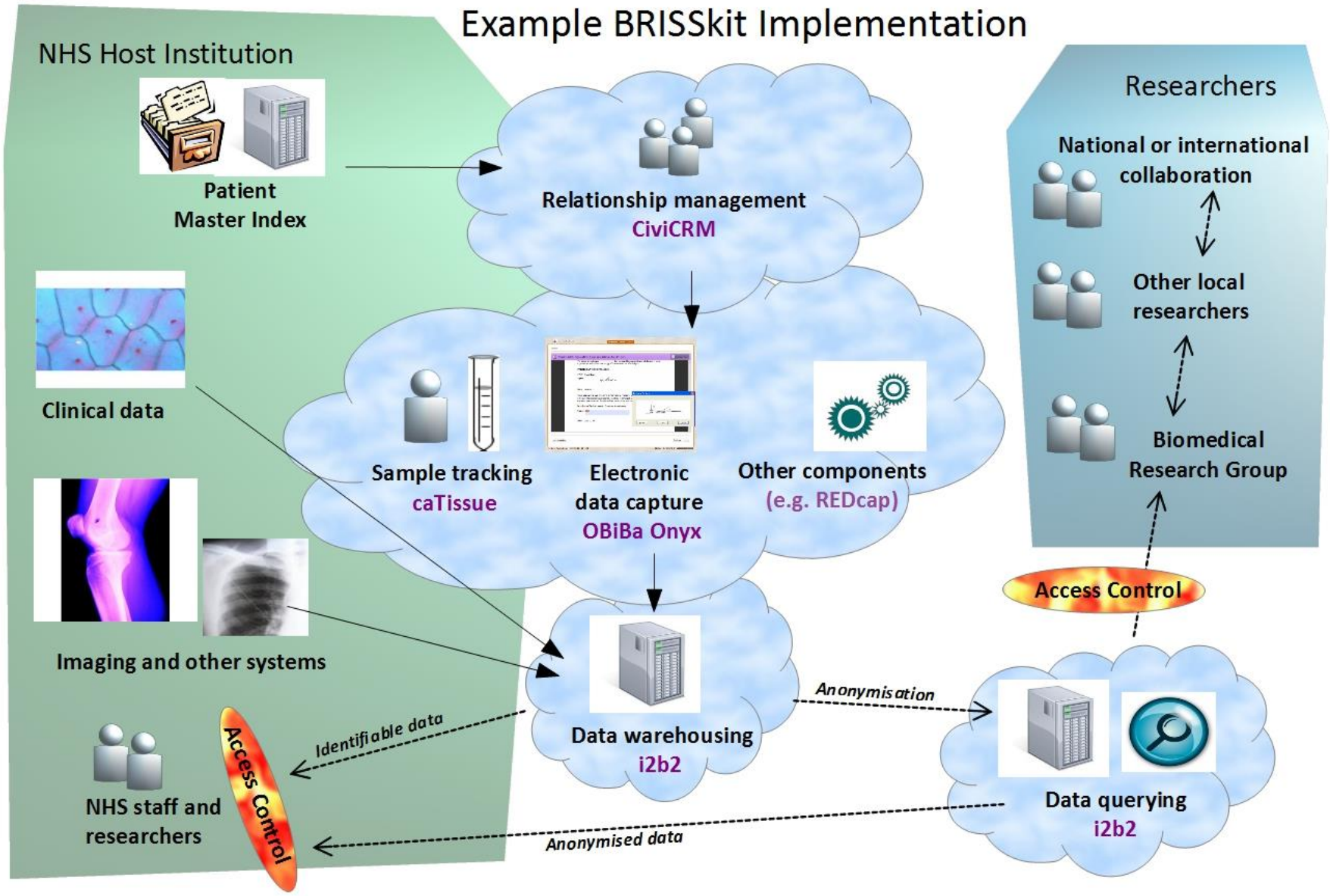
BRISSKit RedCap: survey management



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

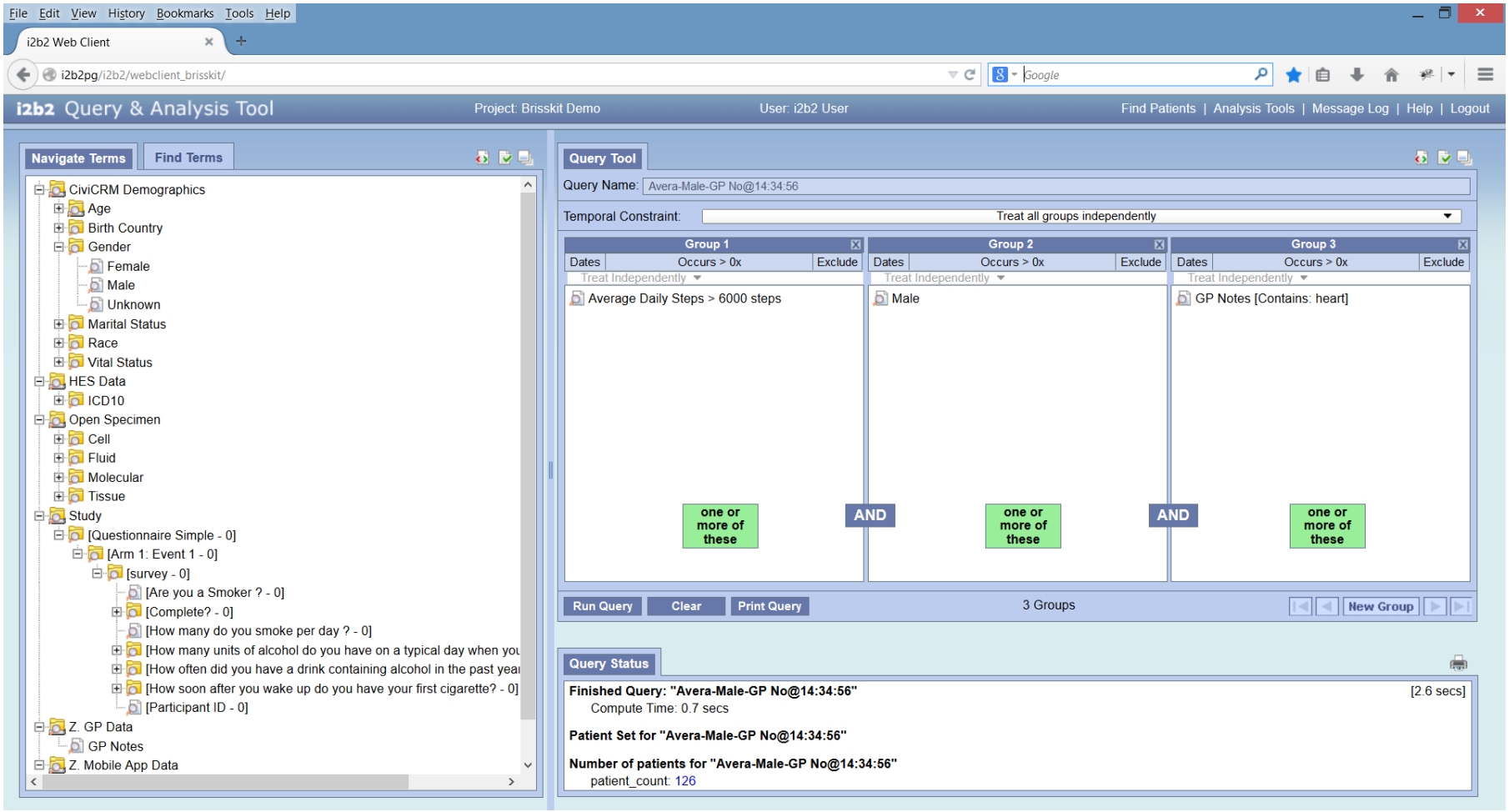
E.g. used for all patient recruits in NIHR UHL Respiratory Biomedical Research Unit – mobile computing on wards and outpatient clinic

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



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

BRISSKit i2b2: data warehousing & querying



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

BRISSKit USPs

- Integrated support for core research processes
- Well-established mature open source applications as protoyped in e.g. 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)
- Direct secure links through Jisc via Janet network

BRISSKit Funding & Partners

- **New HEFCE/Jisc investment** approved for 2014 – 2016
  - Jisc endorsed service
  - Co-design with reorganised Jisc
  - Key Janet Framework partners Farr, Crick, Infinity
- **University of Leicester Cancer Biobank**
  - Tissue sample management built on caTissue, OpenSpecimen
- **NIHR Respiratory Biomedical Research Unit** solutions: University Hospitals Leicester NHS Trust
  - linked to UoL Health Sciences Exceed Study
  - Links to Loughborough-Leicester Lifestyle BRU

BRISSKit highlighted collaborations

- **University of Bristol**
  - ALSPAC Birth Cohort Studies
  - DataShield: simultaneous remote, secure access to multiple large international cohorts
  - SAIL-Farr secure NHS data hosting
- **University Hospitals Leicester NHS Trust**
  - Case Study Module Development
  - UoL Health Sciences Exceed Study
  - NIHR BRUs: Cardiovascular, Respiratory, Lifestyle (Loughborough-Leicester)
  - Leicester Diabetes Centre
- **UoL Data to Knowledge for Practice strategic theme**
  - UoL Genomics, UHL NHS Trust – IBM IT Partnership

BRISSKit Information Governance & Security Management Work Stream - Dr Andrew Burnham leading

1. **Information Governance Toolkit** - analysis of Department of Health (DoH/NHS) IGT requirements vs. BRISSKit 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)

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

BRISSKit Jisc endorsed organisation

Dual model for sustainability proposed (e.g. Ubuntu):

- **.org foundation owns & maintains code**
  - Trustee led
  - Educational
  - Core development
  - Code licensed by not-for-profit
- **.com provides range of service offerings**
  - Modular approaches and scalable tools with open source licenses make good investments
  - Partner with 3rd party technical support e.g. Krishagni (OpenSpecimen)
  - Corporate identity
  - Hosting via Janet/Infinity, SAIL (Farr), private c

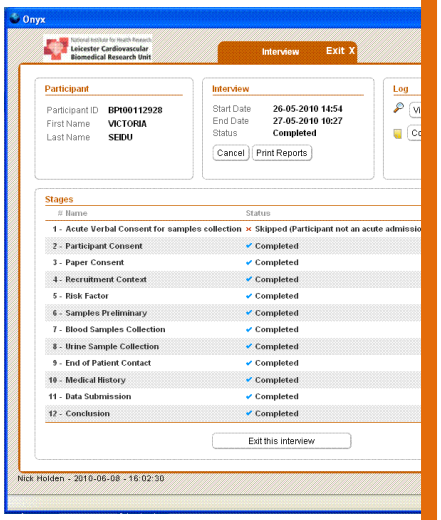
March 2015: i2b2 via public cloud (anon data)

- Users can now get free servers from public clouds, e.g. Amazon, Azure for Research etc
- In March 2015 BRISSKit users will be able to
  - deploy their own i2b2 virtual app onto their own cloud server
  - upload their data through .csv files – default nominal ontology created
  - modify/align this ontology to standardised BioPortal codesets – e.g. SNOMED
  - perform queries on their data using the revised ontology through i2b2

Research: the semantic bridge

Survey dat

Records participa
consent, questio
data and primary
specimen IDs

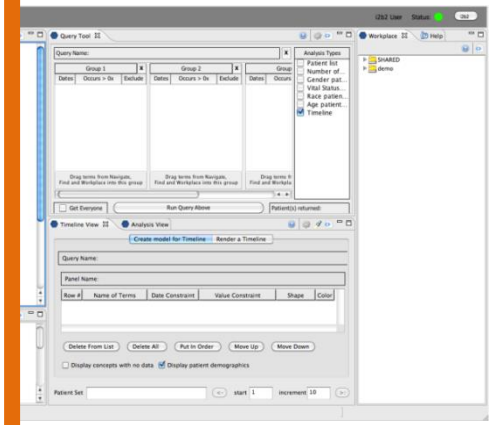


Bio-ontology!

- 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

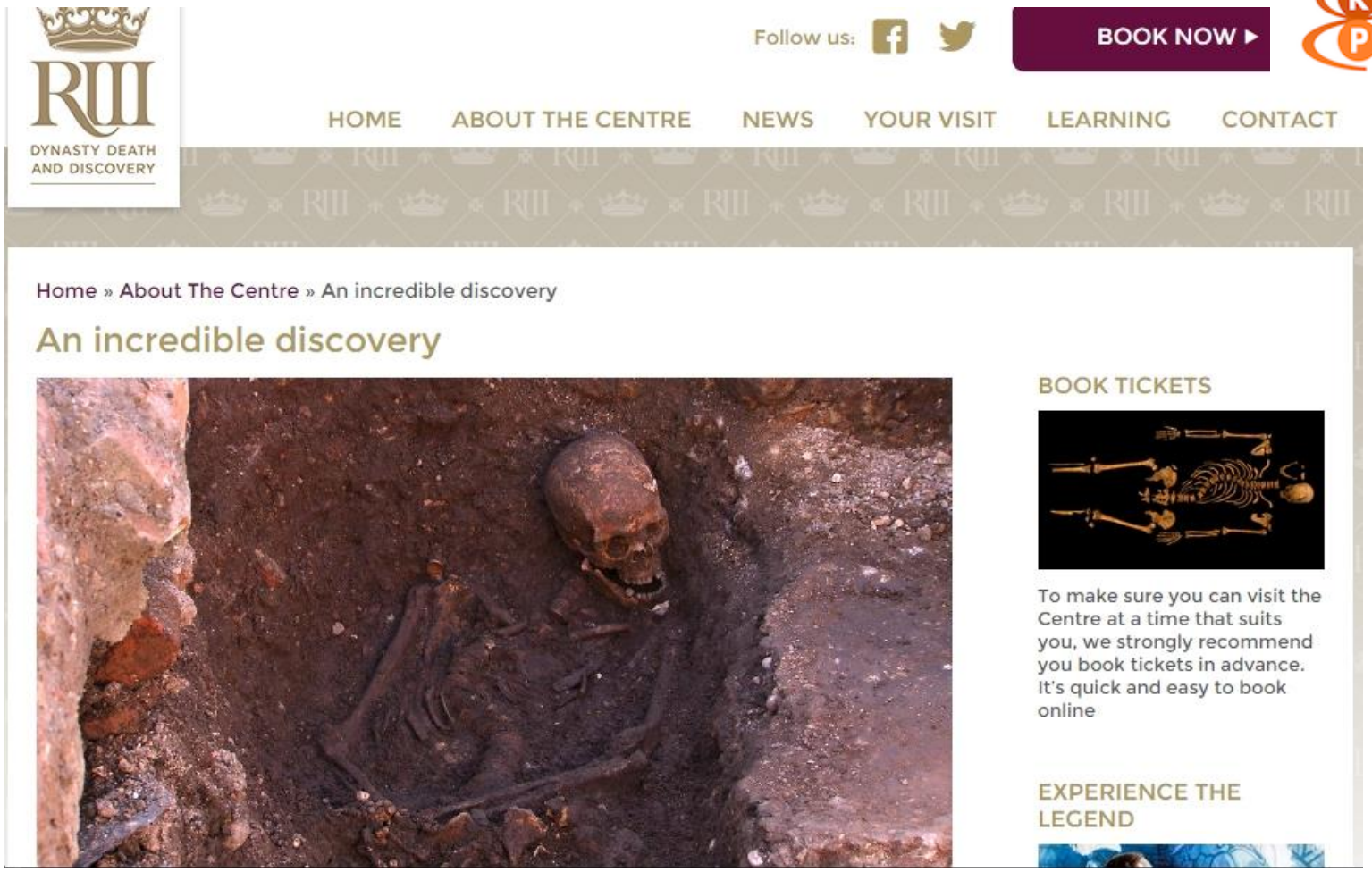
rt selection and
querying



Towards an i2b2 NHS community

- With datasets uploaded into a range of i2b2 instances
- Users will be able to publish their i2b2 datasets
- A community of public cloud i2b2 users will emerge, within which users can publish, exchange and augment data and ontologies
- These merged datasets can then be used to service NHS-wide cohort search, selection and quality management
- Re-identification of cohorts will remain with original sources of i2b2 data

- **BRISSKit Community Event & Health Research Hack: Public Cloud i2b2, 23-24 Feb 2015, College Court, Leicester**
- **Co-located European i2b2 Community User Group Meeting + BRISSKit Leicester June/July 2015!**



DATA to KNOWLEDGE for PRACTICE

University of Leicester

Jisc

BRISSKIT

University Hospitals of Leicester NHS Trust

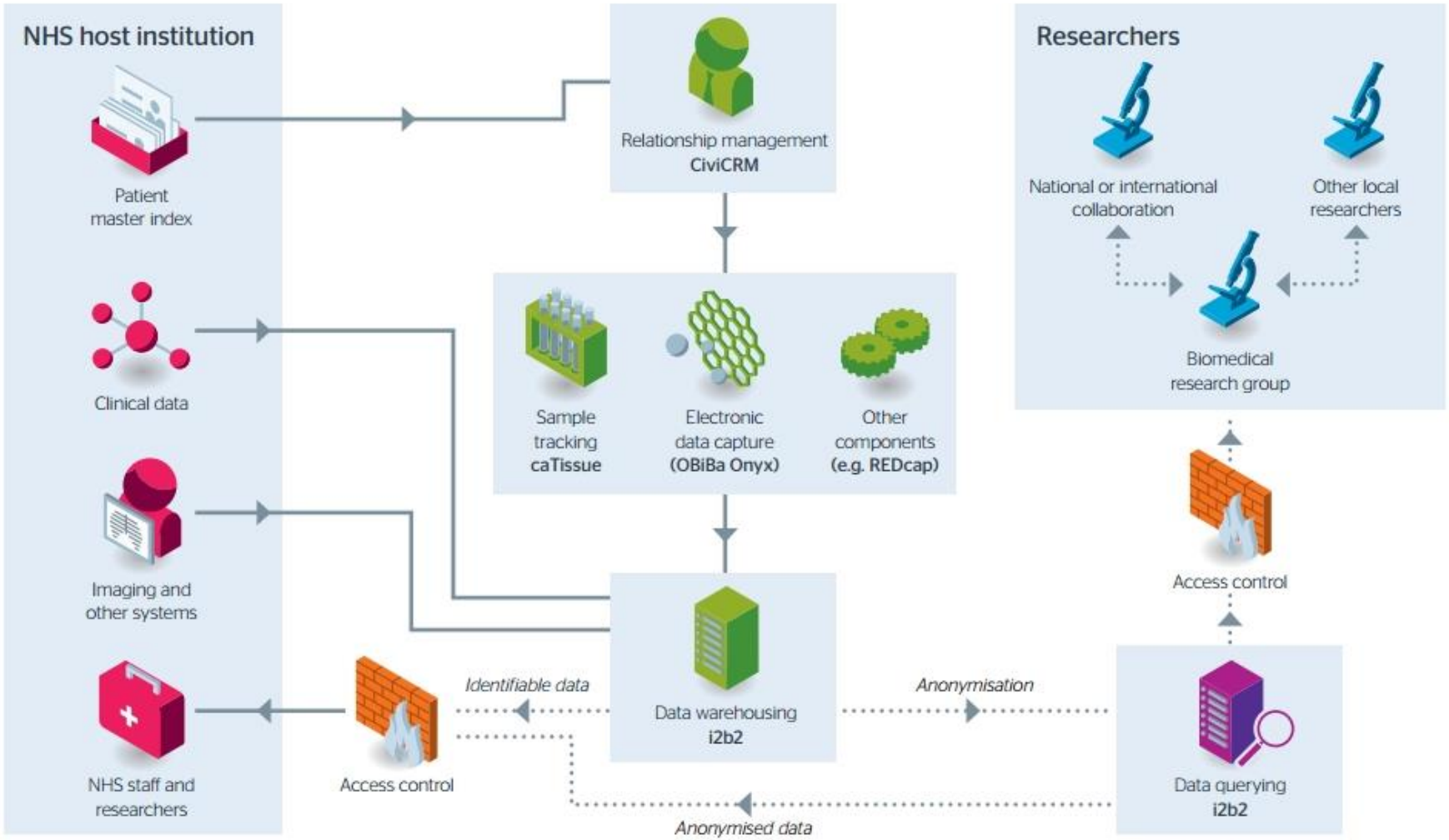
BRISSKit Hack: Public Cloud i2b2 Focus

- 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

Sign up to www.jiscmail.ac.uk/brisskit-announce

Jisc

Biomedical research software as a service BRISSKit A Jisc funded project



Find out more brisskit.le.ac.uk email us at brisskit@le.ac.uk Follow us at @BRISSKit

University of Leicester