University of Leicester

Jisc

NHS National Institute for Health Research

BRISSKit update to NHS-HE Forum, Nov 2013:

Biomedical Research Infrastructure Software Service Kit

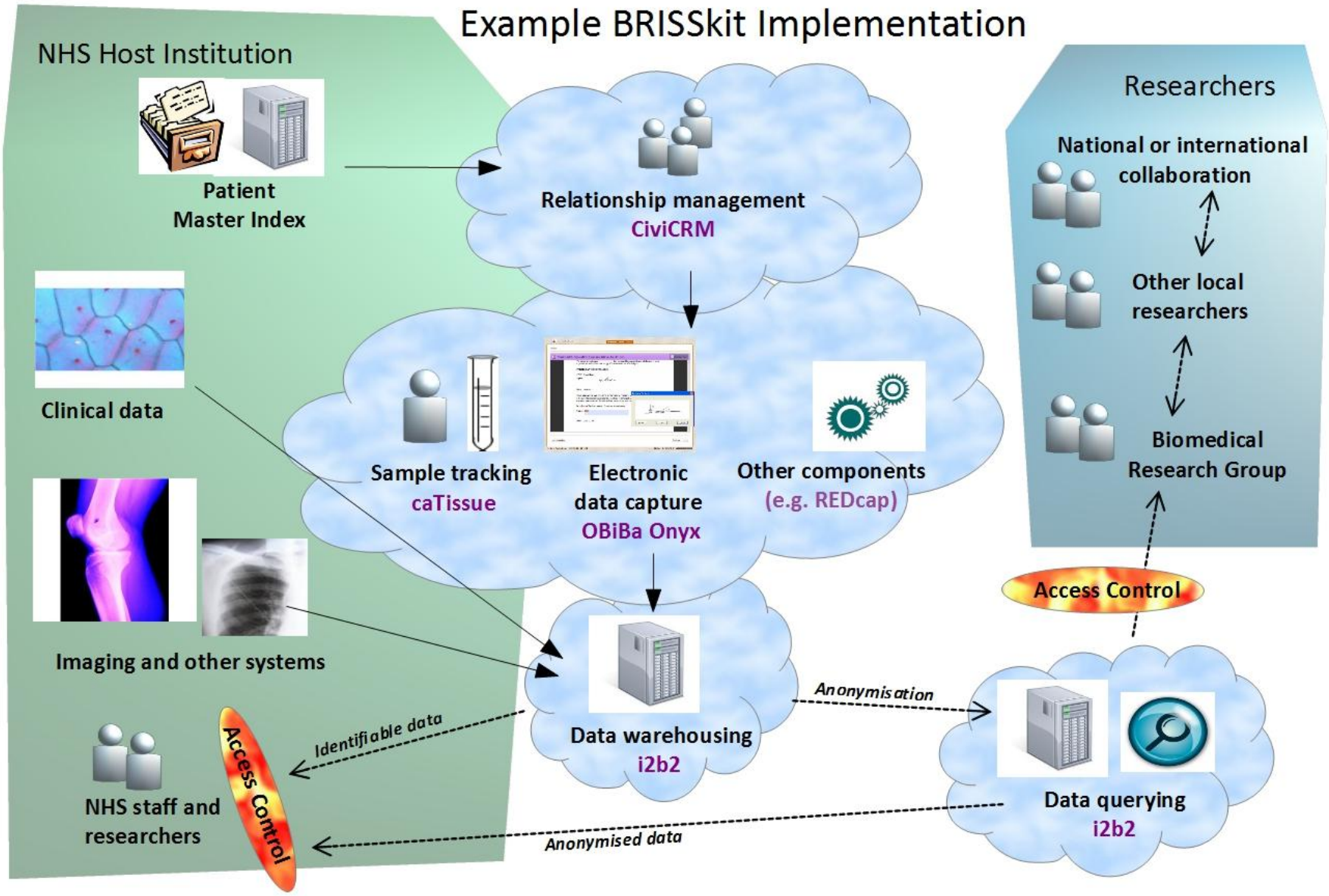
A vision for cloud-based open source research applications

Dr Jonathan Tedds (BRISSKit Project Lead, Senior Research Fellow, Health Sciences)

Health Informatics Group Lead

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

BRISSKIT



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

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)
- Direct links via Janet network / Eduserv dev platform

BRISSKit Funding

- **New HEFCE/Jisc investment** approved for Jan 2014 – Dec 2015 c.£1m+
  - Jisc endorsed service
  - codesign with reorganised Jisc
  - work with Janet Framework partners
- **University of Leicester Cancer Theme Biobank**
  - Tissue sample management built on caTissue
  - Underwritten by UoL Medical College: 2+ years
- **NIHR Respiratory Biomedical Research Unit** solutions: University Hospitals Leicester NHS Trust
  - linked to UoL Health Sciences Exceed Study
  - Links to Loughborough-Leicester Lifestyle BRU & Cancer Theme Biobank

BRISSKit highlighted collaborations

- **University of Bristol**
  - ALSPAC Birth Cohort Studies
  - DataShield: simultaneoues remote, secure access to multiple large international cohorts
  - SAIL-Farr secure NHS data hosting
- **University Hospitals Leicester NHS Trust**
  - 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 Jisc endorsed organisation

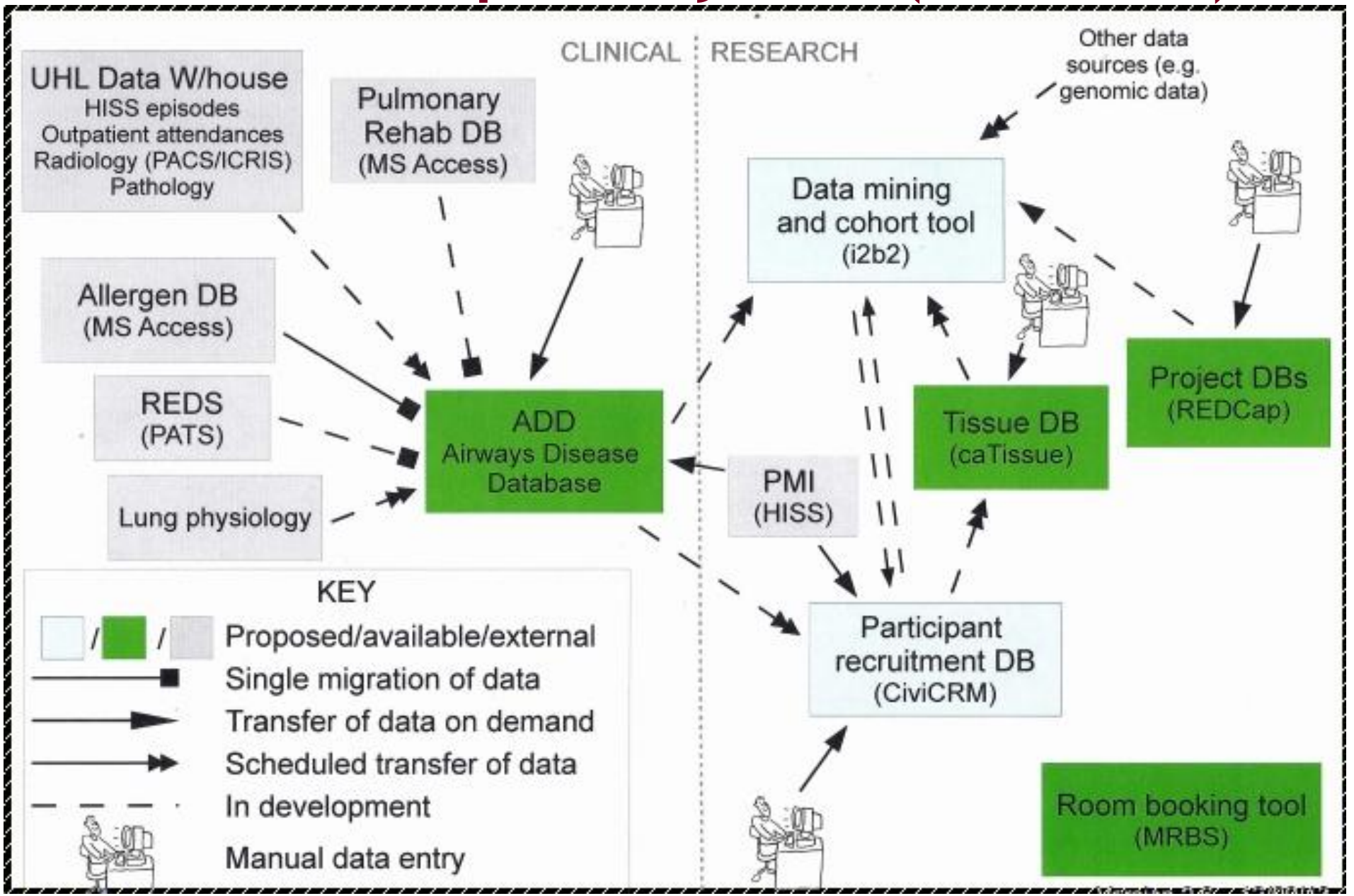
Dual model for sustainability proposed (e.g. Ubuntu):

- **.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
  - Corporate identity
  - Hosting via Janet, SAIL (Farr), private cloud etc

Example implementations: Respiratory BRU

- Incorporate the following BRISSKit components to aid respiratory research :
  - i2b2, caTissue, civicrm
- Configurations :
  - Transfer research specific data from airways disease database into i2b2
  - Provide a flow of caTissue / Redcap data into i2b2
  - Configure civicrm to closely match researchers remit
  - Add Transmart as new component to BRISSKit stack

BRISSKit & Respiratory BRU (Rob Free)



UoL Cancer Theme Biobank UKCBB

- Incorporate the following BRISSKit components to aid cancer research :
  - caTissue, caTissue upload API

- Configurations :
  - Customise caTissue to fit specification.
  - Import existing data into caTissue
  - Provide a mechanism to identify caTissue cohorts between Hospital and University based researchers

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]

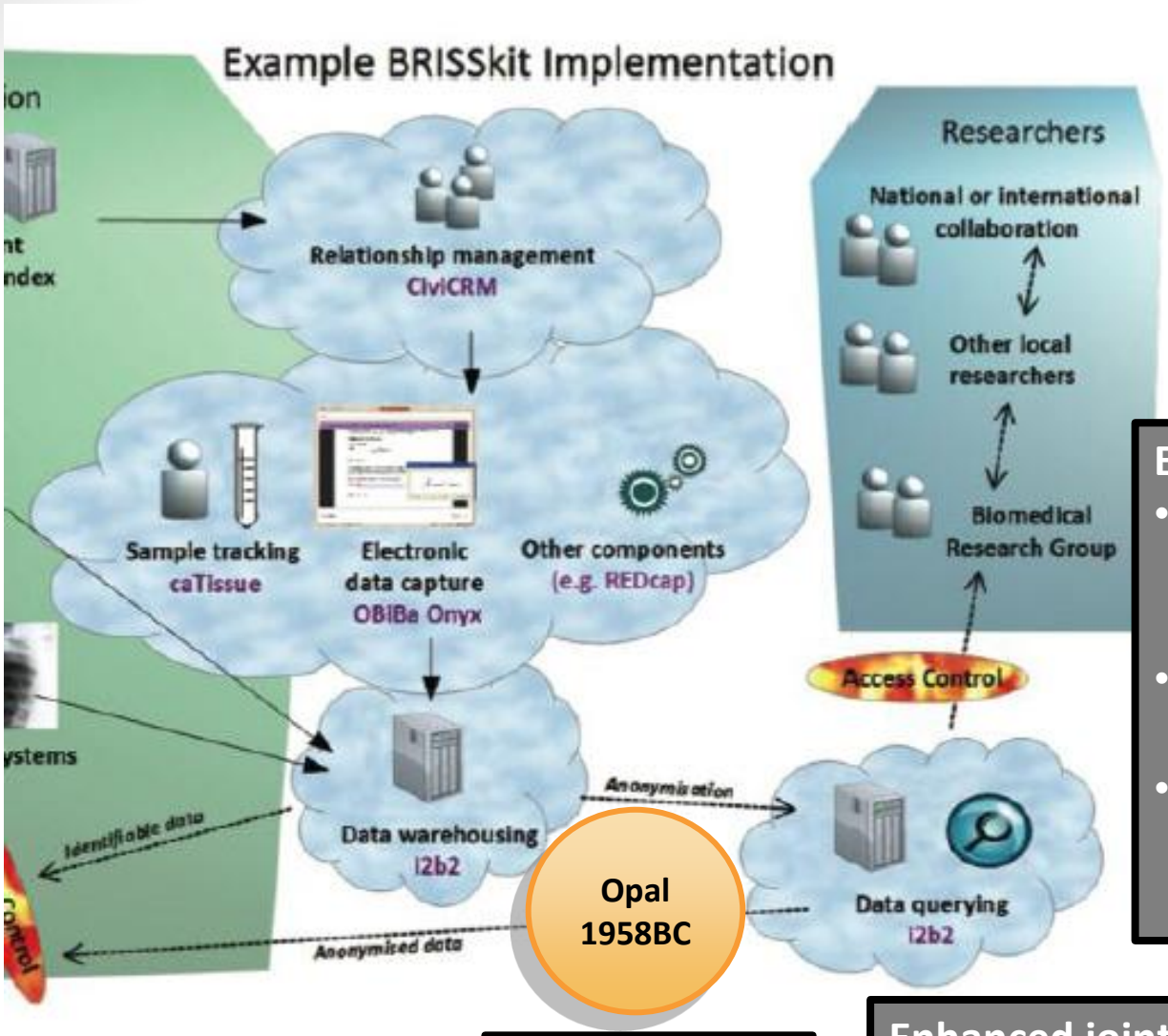
International Journal of Epidemiology 2009;38:263–273
doi:10.1093/ije/dyn147

- Sample size
- Depth of phenotyping
- Quality of measurement

All critical

BioSHaRE.eu

D2K data to knowledge



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

BRISSKit Technical Developments

- API integration platform development key
- Draft timeline
  - Public cloud i2b2 free trial version - July 2014
  - Full NHS secure UoL/UHL BRISSKit instance e.g. UoL Cancer Theme Biobank – Dec 2014
  - Commercial public cloud offer for external research groups – June 2015
  - Secure NHS external hosting e.g. SAIL/Farr implementations + E Mids regional cancer theme biobank – Dec 2015

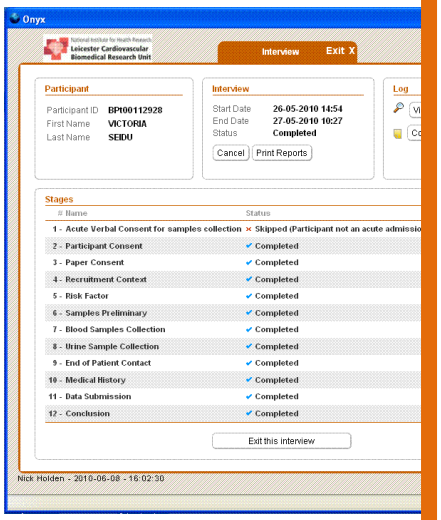
i2b2 via public cloud (anon data)

- Users can now get free servers from public clouds, e.g. Amazon, HP, Azure etc
- In the next phase of 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

OBiBa Ony

Records participa
consent, questio
data and primary
specimen IDs

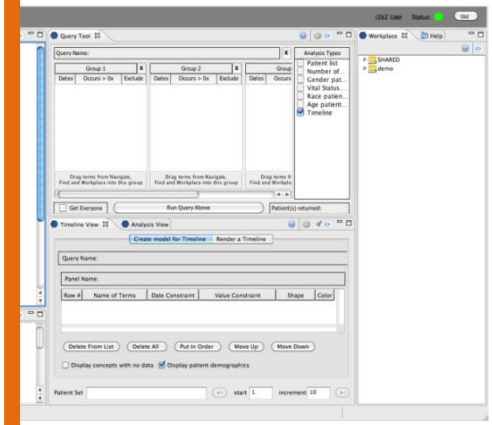


Bio-ontology!

- 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

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

New BRISSKit Community & Hack Events planned for 2014

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

Accepted Research Data Alliance **Interest Group**

Publishing Data

- http://rd-alliance.org/
- **Close coordination with ICSU-WDS working group, CODATA and other ongoing initiatives** in data publication
  - WDS under International Council of Science, RDA wider
  - Avoid duplication within related RDA and WDS WGs – join up
  - For WDS partnerships between publishers and data centres key
- scope the territory – gap analysis
- Use RDA Forum and new http://jiscmail.ac.uk/data-publication 350+ list
- Take findings from RDA / WDS group(s) and trial in other communities / disciplines / institutional repositories

RDA/WDS: Publishing Data Interest Group

Co-chairs: Michael Diepenbroek[1], Eefke Smit[2], Jonathan Tedds[3]

Email: jat26@le.ac.uk. [1]PANGAEA, University of Bremen, Germany [2]STM Publishers, The Netherlands [3]University of Leicester, UK

Introduction

In the empirical sciences, data has traditionally been an integral part of scholarly publishing. In recent decades rapid technical developments, such as digital data and high-throughput techniques, have dramatically altered the scholarly publishing paradigm. This requires new approaches in order to ensure the availability and usability of research data.

Various technical solutions in use or proposed to date offer promise but do not yet provide sufficient benefit and incentives for the data producers themselves and so take up among researchers is still relatively low. The concept of Publishing Data is undergoing a renaissance as part of scholarly communication and on the base of new and proven technologies such as establishing persistent identifiers for datasets. Publishing data offers a strong incentive for researchers to share their data and benefits the wider community through a focus on data quality.

The impact on citation rates is beginning to be demonstrated through bibliometric studies of research articles that include underlying data or are based on secondary reuse of existing datasets such as in astronomy.

The Publishing Data Interest Group brings together all stakeholders involved in publishing research data including researchers, discipline specific and institutional data repositories, academic publishers, funders and service providers. The following 4 initial Working Groups are being developed through the RDA in patnership with the ICSU-WDS and are currently developing Case Statements under the umbrella of the Publishing Data Interest Group. New Working Groups can be formed or join the Interest group as it develops.

Workflows for Archiving and Publishing Data

Jonathan Tedds, Kim Finney, John Helly, Hylke Koers, Fiona Murphy, Amy Nurnberger, Lisa Raymond, Mary Vardigan, Eva Zanzerkia

- Investigate current workflows for archiving and publishing data
- The role of QA/QC and peer-review in the publication process
- The role of science publishers/journals in the data publication process
- Barriers to implementation

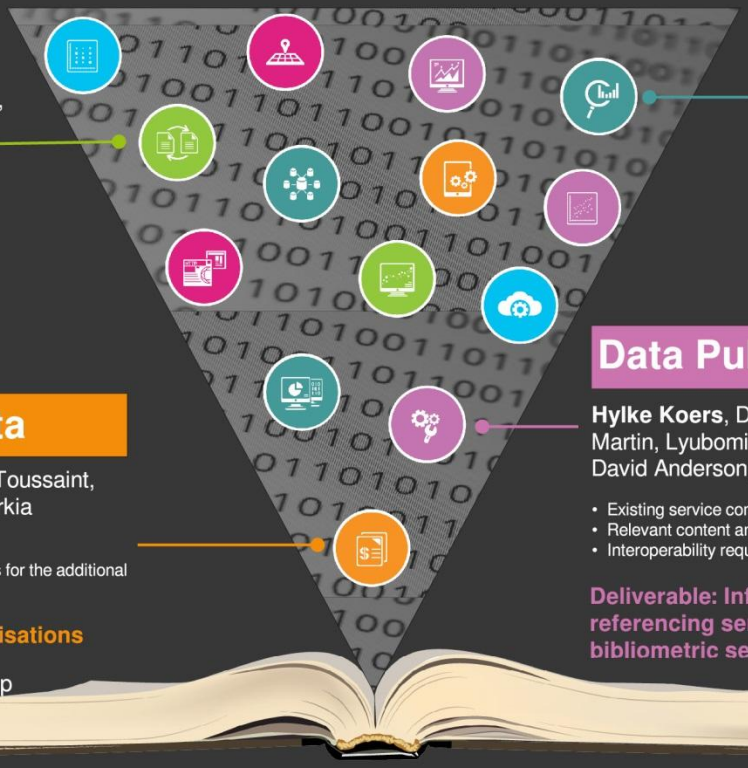
Deliverable: Provide a range of generic and discipline specific workflows for data publication identifying roles, resources and stakeholders

The Costs of Publishing Data

Ingrid Dillo, Simon Hodson, Barbara Sierman, Frank Toussaint, Mark Thorley, Kim Finney, Anita de Waard, Eva Zanzerkia

- Investigate current cost strutures for archiving and publishing data
- Elaborate a business model based on open access which compensates for the additional costs due to data publication

Deliverable: Recommendations for funding organisations



Bibliometrics Including Published Data

Kerstin Lehnert, Euan Adie, Jan Brase, Ross Cameron, Cyndy Chandler, Ingeborg Meijer, Fiona Murphy, Lyubomir Penev, Fiona Nielsen, Nigel Robinson, Mary Vardigan

- General requirements for citability of scientific data (granularity, citation information and persistent identifiction)
- Current citation practice in data centres and literature

Deliverable: Recomendations for data publishers and academic publishers

Data Publication Services

Hylke Koers, David Carlson, John Helly, Francisco Hernandez, Caroline Martin, Lyubomir Penev, Nigel Robinson, Johanna Schwarz, Eva Zanzerkia David Anderson, Juanle Wang

- Existing service components to be used as building blocks
- Relevant content and interoprability standards
- Interoperability requirements for data centres (registration, metadata and data services)

Deliverable: Infrastructure and organisation for a one-for-all cross referencing service for academic publishers and providers of bibliometric services

Scan to visit the Publishing Data Interest Group website: https://rd-alliance.org/internal-groups/publishing-data-ig.html

Scan to join the DATA-PUBLICATION mailing list: https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=DATA-PUBLICATION

www.brisskit.le.ac.uk

Email: *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

D2K data to knowledge

University Hospitals of Leicester NHS NHS Trust

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

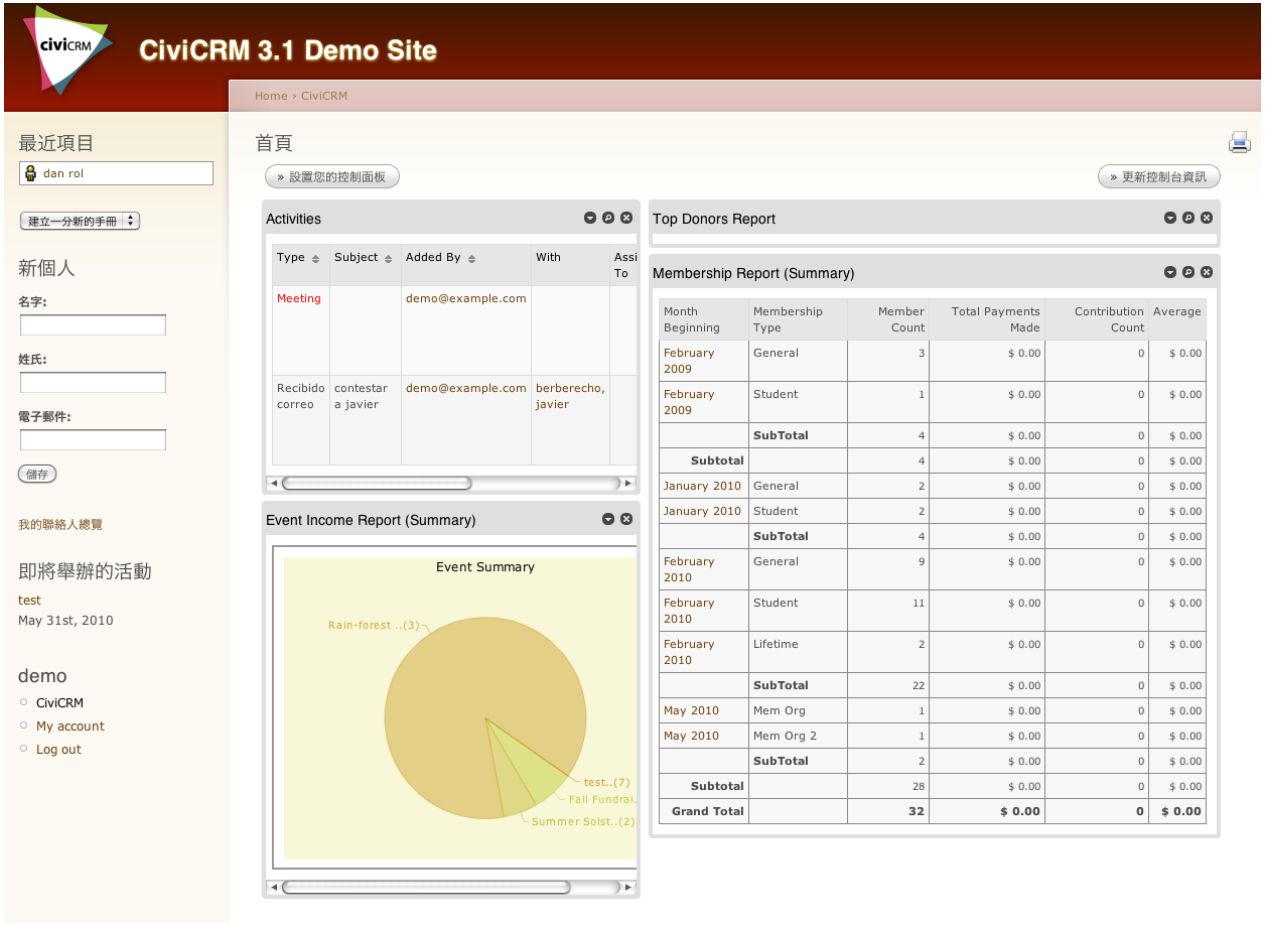
6. Other **NHS Standards/Requirements**:
   - Care Records Guarantee
   - NHS Constitution
   - NHS Records Management
   - Patient Safety DSCN 14/2009, 18/2009

BRISSkit components = web services

CiviCRM

Enables end-to-end contact management for volunteers and research participants, tracking approaches, contact, responses, recruitment, exclusions.

CiviCRM was designed for the 'civic sector' and has an object model that reflects community building and non-profit relationships.

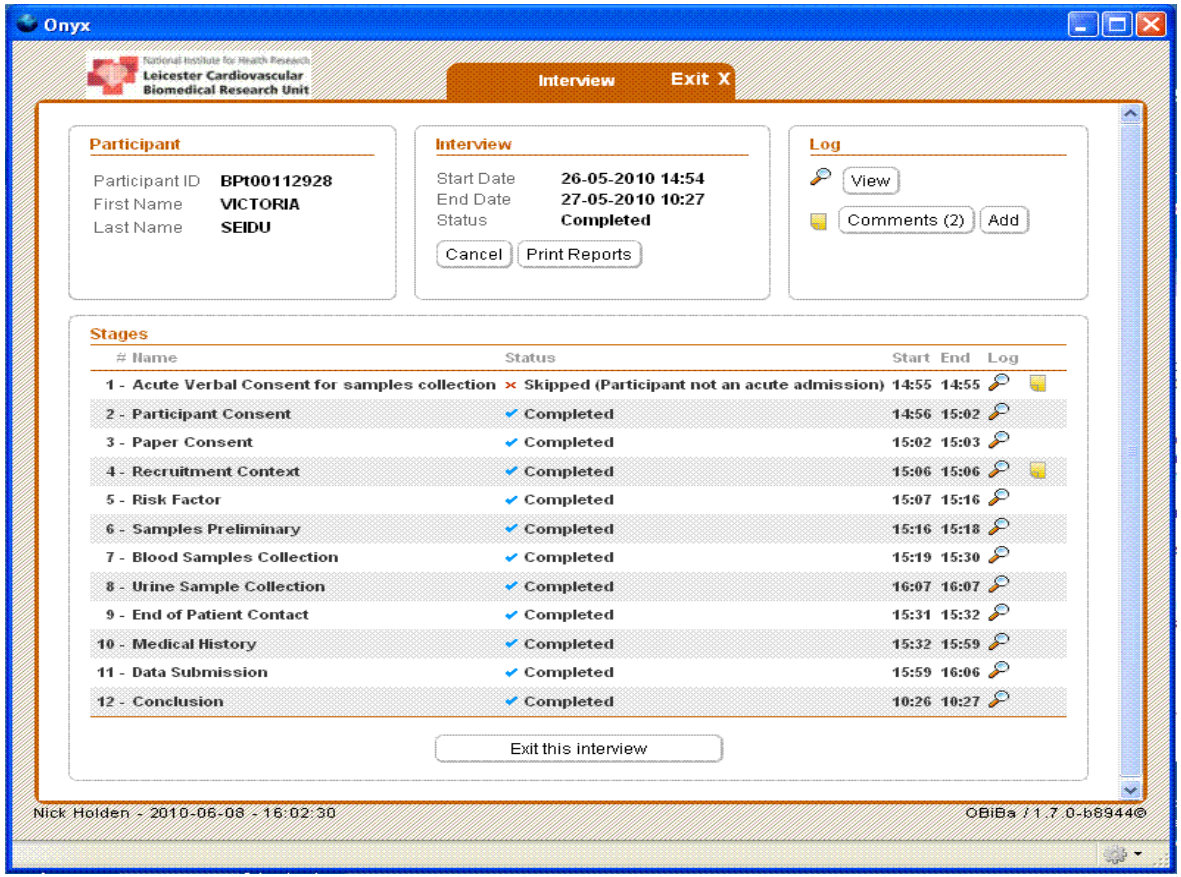


OBiBa Onyx

Records participant consent, questionnaire data and primary specimen IDs.

Web-based, secure data entry by research staff. E.g. used for all patient recruits in LCBRU – mobile computing on wards and outpatient clinic in TMF.

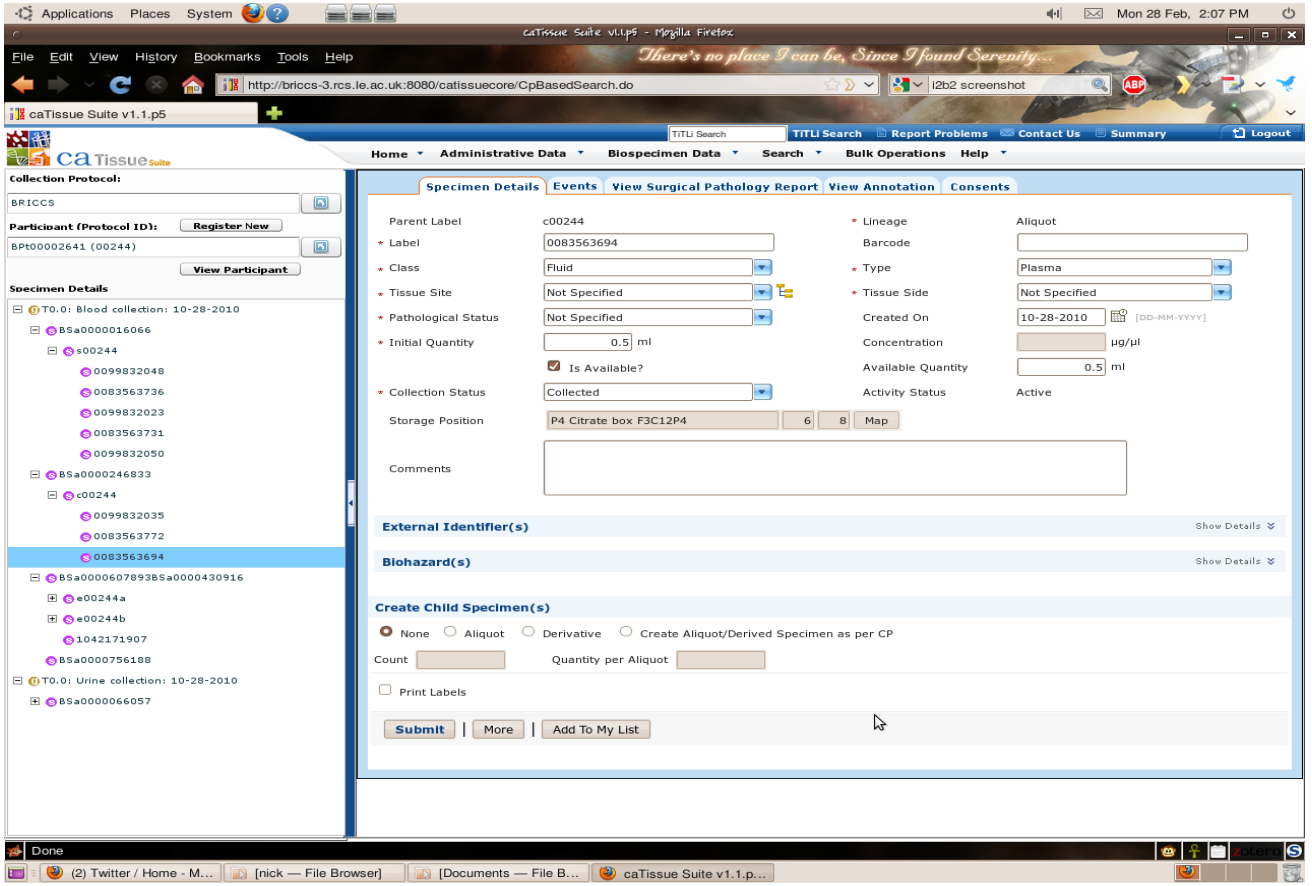
Await significant new release…



caTissue

Holds data on primary, derived and aliquot specimen, including linear and 2d barcodes.

Storage inventory, order tracking – currently over 30,000 LCBRU samples stored and recorded.



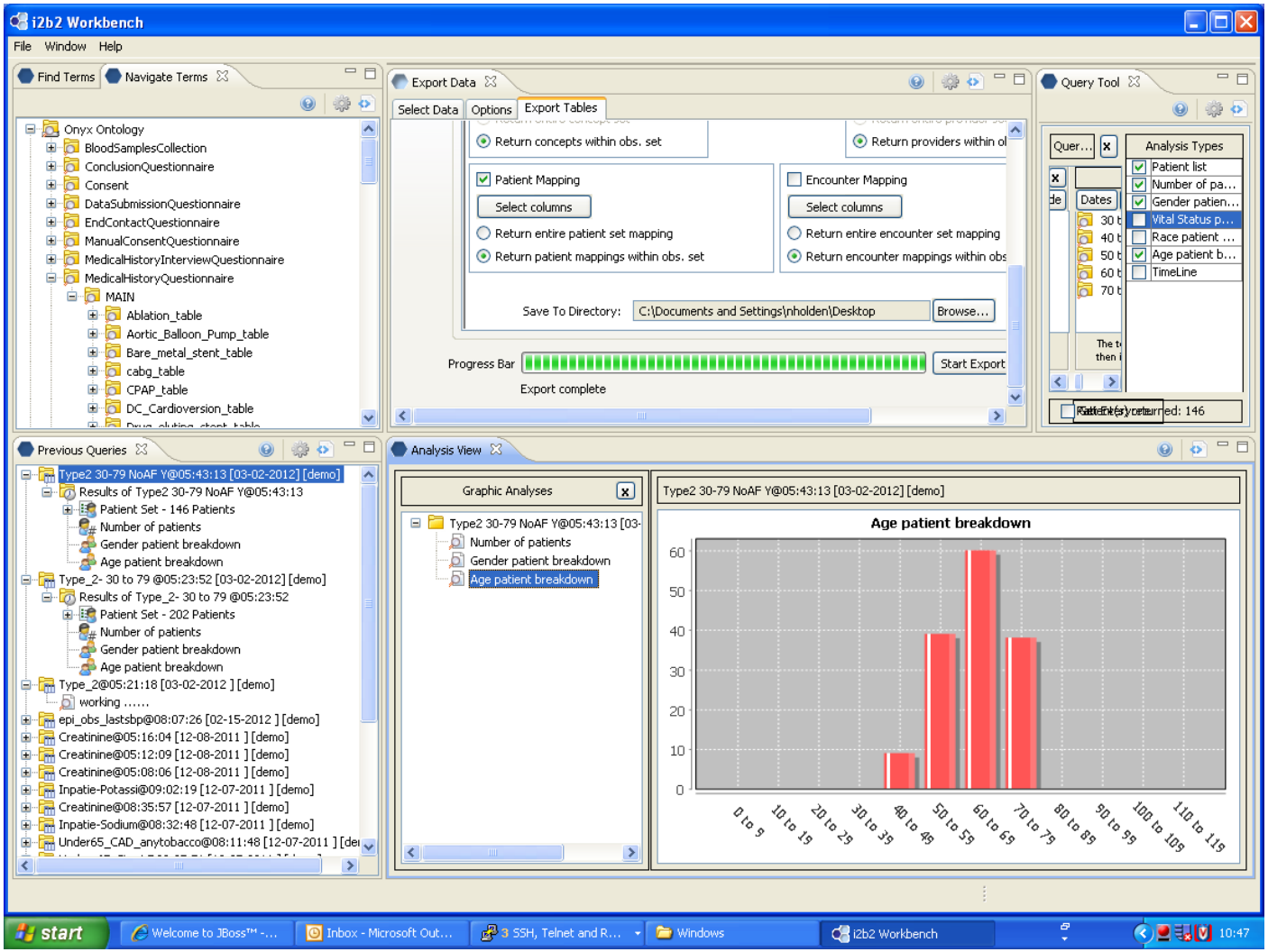
University of Leicester

JISC

University Hospitals of Leicester NHS
NHS Trust

i2b2

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



University of Leicester

JISC

University Hospitals of Leicester NHS
NHS Trust