BRISSKIT

Biomedical Research Infrastructure Software Service Kit

Towards a cloud-enabled translational research IT platform

www.brisskit.le.ac.uk

Introduction

BRISSKit has been set up to help researchers and healthcare workers involved in research. BRISSKit offers a suite of applications to accelerate research by facilitating cohort discovery, making it easier for researchers to manage the identification, selection, engagement and recruitment of suitable subjects for research. It has developed a national shared service brokered by the JANET UK academic network to deliver biomedical research database applications for:

- Contact management and patient recruitment;

- Electronic clinical data capture;

- Tissue sample management;

- Clinical data warehousing.

Aims

BRISSKit aims to:

- Deliver a suite of useful and relevant research applications, as a cloud-based service, accessible at any UK research institution, and meeting the needs of researchers across the fields of biomedicine and translational research;

- Bridge the divide between biomedical research and healthcare by providing secure, reliable and affordable open source database solutions provided in-house or as a national cloud-based service via the UK JANET academic network or secure NHS network provider. The solutions will support the management and integration of tissue samples with clinical data and electronic patient records, having appropriate information governance and anonymisation where required;

- Accelerate translational research in the UK by providing a software service available on a low cost pay-as-you-go pricing model that would otherwise cost many tens or hundreds of thousands of pounds to replicate within a research institution;

- Deliver state-of-the-art functionality with a robust and reliable service model, ensuring researchers meet their information governance, data protection and confidentiality requirements without sacrificing research objectives;

- Facilitate the use of open standards, open source software and the sharing of research data and workflows, allowing collaboration by the widespread use of common tools and the appropriate use of data sharing and anonymisation technologies.

The service is unique in offering:

- Fully UK customised versions of mature open source applications used more widely internationally: CiviCRM, Onyx, i2b2 and caTissue;

- Seamless integration of applications through well-defined use cases;

- Easy set-up through browser access to cloud configuration, capable of being hosted in any compliant cloud provider;

- API and integration tools available for integration with clinical systems or your own tools.

The service can be tailored to fit the needs of many research groups and projects and promises significant savings in research database and IT support costs.

Research groups

BRISSKit (via its use of the i2b2 query tool) allows highly complex criteria to be applied to populations. A description of a potential cohort might be along the following lines: "I'm interested in finding Asian and matched Caucasian gentlemen, aged between 18 and 55, suffering from type 2 diabetes, who are within a normal range BMI, regular smokers, have an elevated HDL level and have an available DNA sample to genotype prior to participant
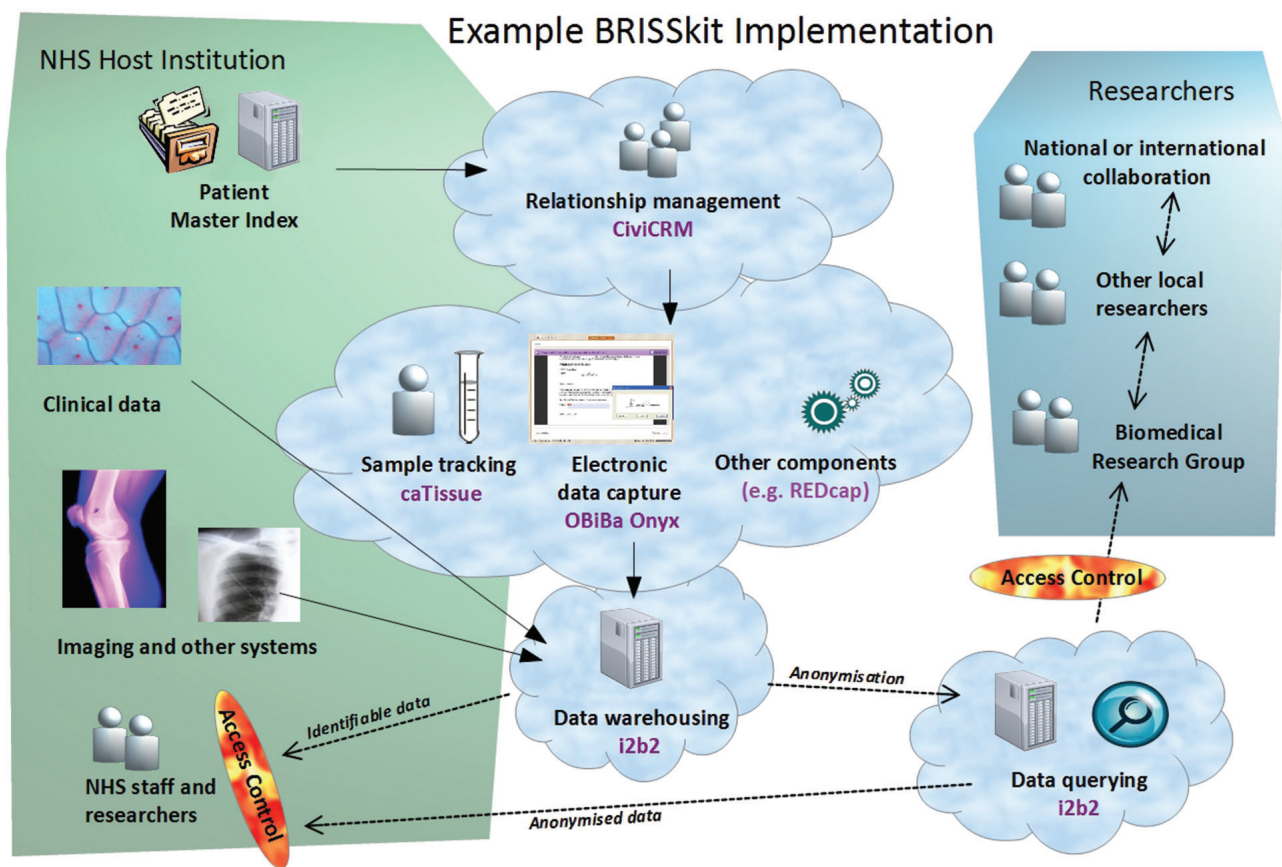
contact". Once a cohort has been selected along these lines, it can be returned to CiviCRM for re-identification, to manage further patient contact for possible clinical trials, or for inducting into additional detailed research interviews and tests.

With advances in research practice, and a rapid increase in the scale of clinical and genomic data available to translational researchers, the need for informatics tools to support their practice is now overwhelming. Typically, larger research groups either build their own or procure and tailor expensive commercial solutions while hundreds of smaller research groups rely on ad hoc spreadsheets or paper records only. BRISSKit offers the potential for a dedicated research group database, ultimately linked to clinical data repositories, blood and DNA samples, genomic data and knowledge extraction from documents and image stores.

Research data will be more easily and securely managed using affordable community cloud-based infrastructure, not only allowing for sharing of data between researchers within an institution but also enabling multi-site aggregation of cohorts where required. Components will be provided based on open source solutions that can be delivered as a service from a secure community cloud or hosted locally as appropriate, and accessed securely using common web browsers. Ultimately, BRISSKit becomes a platform for additional tools and integration techniques to allow data to move directly from clinical applications to research.

Example BRISSKit Implementation

NIHR Leicester Cardiovascular BRU

The NIHR Leicester Cardiovascular BRU at Glenfield Hospital is part of a multimillion pound scheme to prevent, diagnose and treat ill health. It is one of 16 BRUs around England funded by the National Institute for Health Research (NIHR). The Glenfield Unit is a partnership between the University of Leicester and University Hospitals Leicester (UHL).

The NIHR Leicester Cardiovascular BRU is specifically aimed at improving the diagnosis, prognosis and treatment of coronary heart disease and hypertension (high blood pressure). It provides an infrastructure – people, space and equipment – that allows researchers to complete challenging or complex projects in these conditions.

In addition, the BRU hosts a dedicated research database, the first of its kind in the UK, which is linked to a collection of blood samples from patient volunteers who have made their medical records available to researchers.

Background
Undertaking translational research requires extensive use of computer software to collate, record, integrate and analyse data. The NIHR Leicester Cardiovascular BRU first embarked on the widespread use of Free and Open Source Software (FOSS) in translational research to develop the Biomedical Research Informatics Centre for Cardiovascular Science (BRICCS) to provide end-to-end informatics support for the research process.

Deploying an array of open source applications to support clinical and translational research is unusual in the NHS, but more common elsewhere in the world. FOSS promotes the use of open standards for data interchange, increases the potential for national and international collaboration, and offers significant savings.

BRISSKit is taking this a step further and seeks to share the benefits of BRICCS by providing easily configurable cloud-hosted software as a modular service. It will adopt the existing open source applications that have been further developed, tested and used within the NIHR Leicester Cardiovascular BRU by the BRICCS project. These shared services will be hosted in the cloud at a national scale over the secure JANET academic network or NHS provider with appropriate information assurance.

Institutional benefits:
- Flexible web services via a browser that can scale to meet demand;
- Trustworthy cloud solutions brokered by JANET to include NHS N3 for sensitive data;
- Lower costs, with no frontloading of cost;
- Reliability and stability in the face of hardware failure;
- Reduces the scaling cost of computing to projects;
- Lends itself to distributed projects and teams;
- Integration of distributed infrastructure;
- Enhances the basis for data sharing, reuse and federation;
- Ensures sustainability of the service;
- Enables solutions for confidentiality

and NHS information governance compliance;

- Bridges the gap between NHS healthcare and university domains.

Researcher benefits:
- Integrated and mature open source IT infrastructure for biomedical research;

- API and integration tools available for clinical systems integration;

- Anonymised versions of patient-sensitive databases are hosted for university researchers;

- Collect data from consented participants and combine with NHS data;

- Sophisticated cohort selection criteria for research;

- Affordable alternative to build-your-own solutions or expensive procurement;

- Partnering opportunities;

- Benefits of BRISSKit securely managed cloud-based or local services;

- Sustainability solutions.

National strategy benefits:
- BRISSKit is fully aligned with the Cabinet Office ICT Futures strategy, which calls for all development projects to consider open source software and open standards;

- It supports the national collaboration for growth agenda including Cabinet Office/NHS open data measures by facilitating data harmonisation, data sharing and collaboration nationally and globally.

The Biomedical Research Informatics Centre for Cardiovascular Science (BRICCS)
BRISSKit builds on the tools already developed for the BRICCS project. BRICCS is a system for capturing and coding clinical information about patients with cardiovascular disease seen in UHL specifically for research purposes and for facilitating recruitment of patients into new studies. All patients attending relevant departments in UHL are considered as potential research participants and their consent is proactively sought to access

their clinical notes and data for research. UHL already collects a large amount of information on patients, both of a generic nature and more detailed information on specific groups for national audit, which are of interest to the NIHR Leicester Cardiovascular BRU, e.g. those undergoing cardiac catheter-lab related procedures or coronary artery bypass and patients with myocardial infarction and acute coronary syndromes.

BRICCS is a software platform to support translational research within the BRU. It will save precious time for new research by leveraging existing clinical data for research purposes.

The core of BRICCS is a dedicated research database, the first of its kind in the UK, linked to a collection of blood samples from patient volunteers who have consented to make their medical records available to researchers. After nearly 2 years of operation, BRICCS has recruited over 3,000 patients and has over 85,000 samples stored in the biobank. This enables us to match suitable patients to particular future studies, and speed up the recruitment phase of new research.
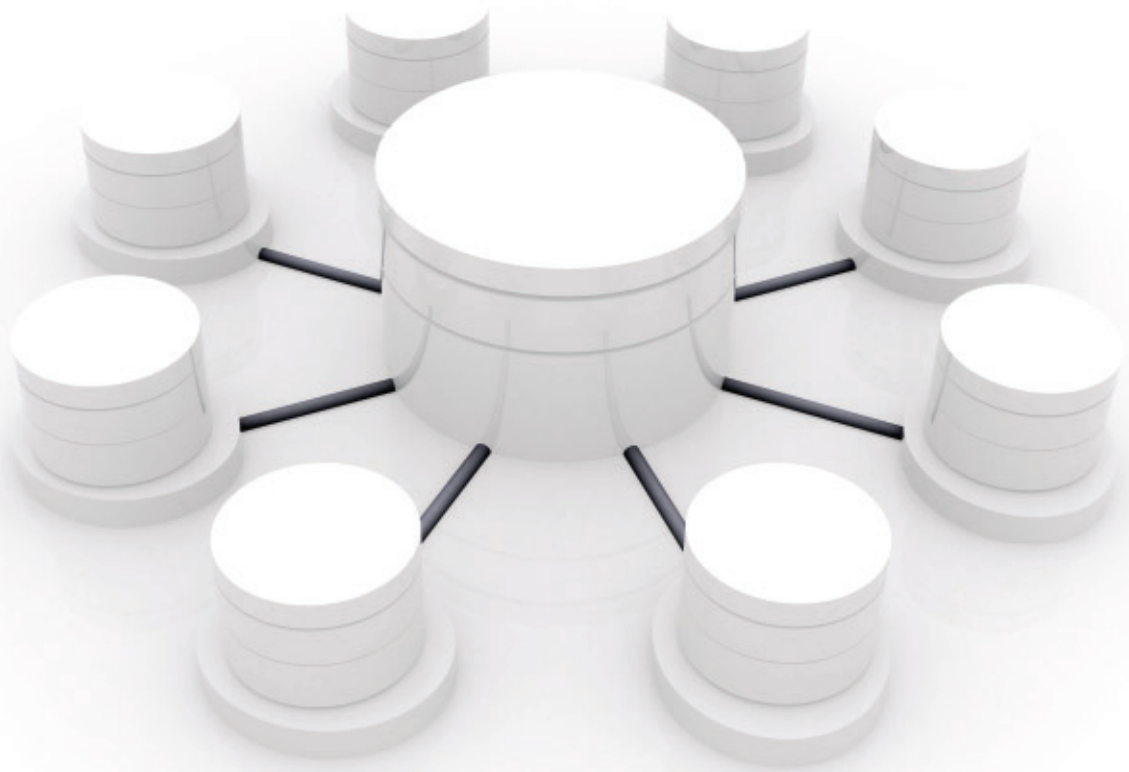
What is open source software?

All BRISSKit software is open source software (OSS). Software is copyright material, and virtually all software comes with a licence defining what you can do with it. Open source software is software that has been released under a licence and meets the criteria of the open source definition, described in full at:
www.opensource.org/docs/definition.php

Open source software is not necessarily free of cost, and is not the same thing as 'freeware'. BRISSKit uses free open source software. Open source software must be subject to an OSI approved licence:
www.opensource.org/licenses/alphabetical

OSS licences do differ, but generally allow royalty-free usage. These licences allow modifications and derived works to be distributed under the same terms, and encourage contribution back to the core libraries. OSS licences also encourage interoperability with either proprietary or other open source software and hardware platforms and can be supported by any organisation with the right knowledge and skill with the source code.

The commercial and business world has embraced OSS following the realisation that it makes commercial sense. Any organisations that need software to perform a particular function can either develop it in-house (or contract it out), in which case they bear the full cost of development and support, or they can purchase proprietary software, which invariably requires expensive customisation or extension to suit their specific needs. OSS allows them to share the costs and risks.

OSS is particularly suited to developing solutions in niche areas, including scientific research, where the community of researchers in a particular field may be small. Proprietary software is not cost-effective in developing such specialised software. Open source is the best way for scientists to share the cost of developing software to support their research.

OSS sustainability relies on the community of contributors taking on a number of roles, related to writing and updating code, assisting users, testing, investigating bugs, writing documentation, translation, promotion and governance. A governance model lays out the ground rules for participation and the processes for communicating and sharing within the project team and with the community of contributors who are vital to sustainability.

What is cloud computing?

Cloud computing refers to the development and use of software hosted on servers owned by external service providers. The service provider ensures service levels that are consistently higher than in-house provision, including software availability, performance, backup and data security, so the management overhead to the customer is minimised. However, the owner of the data entered into the cloud-hosted application retains ownership of it, and access is allowed only to owner-authorised users.

Access to the software requires an internet connection, and access through a viewer or internet browser. Several models of cloud computing exist to support different modes of customer access: applications (as a service), platforms (as a service) and infrastructures (as a service).

BRISSKit offers an integrated application as a service model, where all applications are available either individually or as a fully integrated application set.
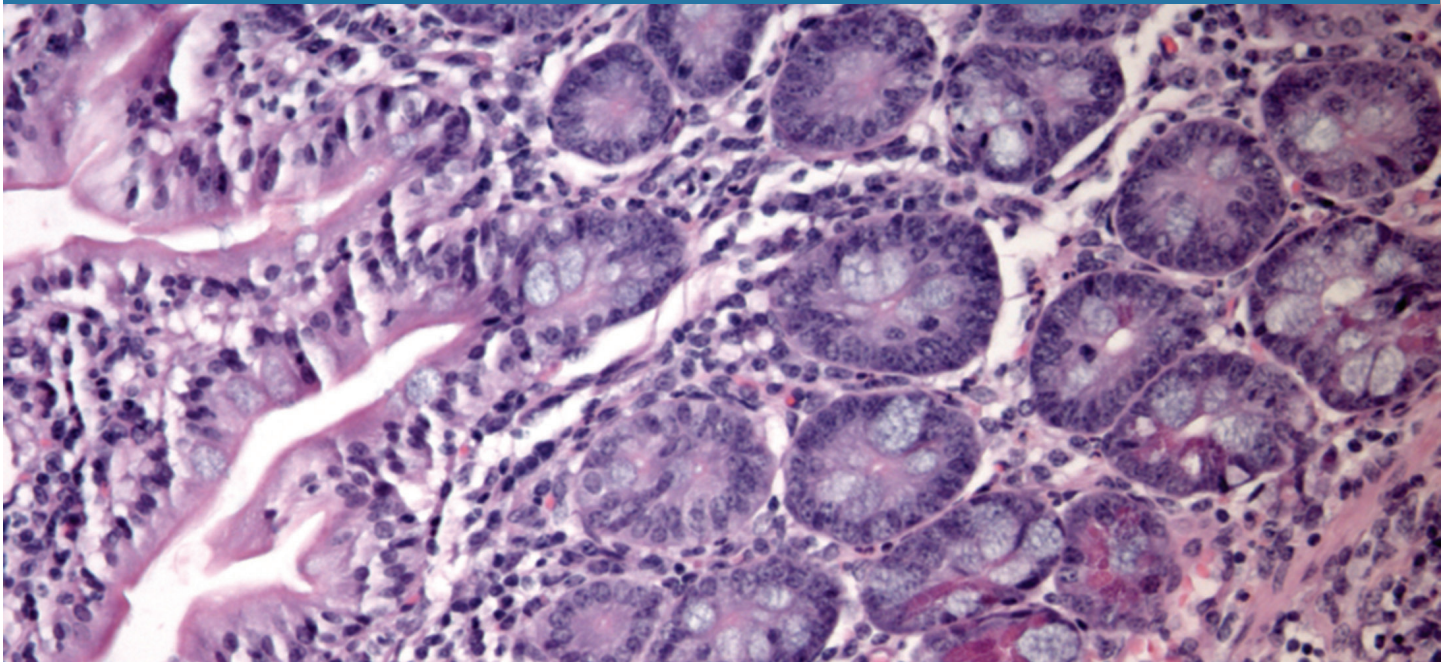
BRISSKit community feedback

The development of BRISSKit has only been possible through bringing together a community of scholars, data scientists and research IT specialists from a range of groups in the University of Leicester, with the support of the Pro Vice Chancellor for Research & Enterprise, including the Departments of Cardiovascular Science, Genetics, Health Sciences and Physics & Astronomy alongside the University IT Research Computing Service and NHS Trust IT support service. Nationally, the project has drawn upon expertise in the HEFCE/JISC University Modernisation Fund including Eduserv, the JISC Managing Research Data Programme, the National Institute for Health Research and the UK academic network JANET Brokerage programme.

National workshops in 2012 – held for representatives of biomedical research groups and institutions – highlighted the benefits of the BRISSKit approach and laid the foundations for ongoing collaborations including wiki to support the BRISSKit community, which is now provided via www.brisskit.le.ac.uk.

Most research institutions are in favour of open source in principle. Those already using it are enthusiastic and would want to adopt a 'menu' approach to BRISSKit where they could select particular elements. Institutions not yet familiar with open source are likely to be convinced by user endorsements. The ever-rising licence costs for commercial solutions are a deterrent, which lead to vendor lock in and may prove to be a poor fit. On balance, most participants would be willing to contribute

towards the costs of ongoing support or customisation, particularly if satisfied that specific features such as audit trails for ethical approval are included.

Institutions that do not have access to IT expertise and infrastructure of their own are most likely to welcome a cloud solution, subject to reassurances with regard to governance, NHS N3 accreditation, data security and shared access to patient data. A free trial period would be favoured in order to demonstrate that BRISSKit will integrate with legacy data and existing tools.

Within the NHS there may be some resistance to moving from established commercial suppliers, but national incentives and changes in the funding and licensing climate may encourage greater openness to new solutions. Some researchers and institutional staff may be reluctant to move from platforms that they are familiar with, or may see the cloud as losing control and ownership of data held locally, so a facility to withdraw data at any time would provide reassurance.

Validation, with flexibility in the application of standards and ontologies was mentioned as one area where tools could make a useful contribution.

Key benefits that the community have highlighted include:
- Savings in hardware costs, software licensing and staff costs;

- Speed of deployment of tools;

- Reduced management overheads;

- A cloud solution means data from a growing number of studies can be made available in one place and accessible from everywhere;

- Facilitates collaboration and use of a common language;

- Would make data from the NHS available to the research community and vice versa;

- A product developed by the research community for the research community.

BRISSKit status

During 2012, the BRISSKit project team have deployed standalone and integrated BRISSKit applications over the UK JANET academic network in an Eduserv hosted data centre. This enables us to provide early adopter testing for our pilot user groups in UHL including cardiovascular, respiratory and lifestyle research areas funded through NIHR. We are also supporting pilot groups at the Institute of Child Health at University College London and the School of Cancer Studies, University of Birmingham. BRISSKit has been extended to support greater integration of data across the applications, and options for hosting are being evaluated that allow BRISSKit to be used with NHS data in accordance with NHS information governance recommended best practice. Further work will also be done to create a simple method of deployment of BRISSKit within user institutions and to facilitate private cloud configurations.

More information
Dr Jonathan Tedds – Project Lead
brisskit@le.ac.uk
www.brisskit.le.ac.uk

BRISSKit ingredients

CiviCRM
www.civicrm.org

CiviCRM is an open source Constituent Relationship Management (CRM) solution. CiviCRM is web-based and was designed specifically to meet the needs of non-profit, advocacy and non-governmental groups. It has a hugely flexible information model, allowing for the registration and recording of individuals, families, groups, organisations, their complex relationships and interactions between the organisation running CiviCRM and their clients, volunteers, customers, donors, employees, vendors, activists, etc. Contact with individuals or groups can be in the form of circulars, mailshots, phone calls, meetings, appointments, interviews or any other form of engagement that can be added to CiviCRM's records.

BRISSKit exploits this powerful flexibility to offer research teams or institutions a mechanism for managing the dialogue with research volunteers from the very moment

that they are identified. Whether they are identified for inclusion in studies by clinical colleagues or volunteer to take part in research on their own initiative, CiviCRM can record the details, the purpose and the outcome of each contact, including noting where and when volunteers are considered for inclusion in a specific research project, and whether they meet the relevant inclusion/exclusion criteria. As a result, CiviCRM will deliver statistics on recruitment processes for individual studies or across the whole research institution.

CiviCRM is in use around the world in a vast number of different configurations for organisations ranging from Amnesty International to the Free Software Foundation. CiviCRM was SourceForge's 'project of the month' for January 2011.

With the BRISSKit configuration of CiviCRM, when a volunteer is registered onto a

particular study, records can be created for them within the data collection, specimen inventory and clinical data warehouse applications automatically. CiviCRM can also help ensure that volunteers are not repeatedly contacted by researchers running related studies, and that those who are recruited to a study are not approached for other involvement until a suitable time.

CiviCRM benefits the researchers by making their processes more efficient, the research managers by making their activity more accountable, and the research subjects themselves by making their involvement more effective.

OBiBa

OBiBa Onyx
www.obiba.org

OBiBa is a collaborative international project that builds high-quality open source software for biobanks. OBiBa is a core project of the Population Project in Genomics Consortium (P3G), an international organisation fostering collaboration and knowledge sharing for population genomic studies.

Onyx was developed to collect data from 300,000 participants for the Canadian Partnership for Tomorrow (CPT). Onyx is a web application that centralises and manages participant baseline interviews at assessment centres or clinics. An interview is represented as a set of interdependent stages that participants have to go through in order to complete their visit (consent form signature, health questionnaire, biosample collection, and so on). Onyx stores the data collected during the stages centrally and makes it available to all workstations.

Onyx is used in BRISSKit to provide a questionnaire builder for observational research projects, which can include data collection from physical measurement devices (see OBiBa website for details). Participants can be submitted to Onyx from appointment lists, as 'walk-in volunteers' or by reference to external systems such as hospital patients indices. At the NIHR Leicester Cardiovascular BRU, the Onyx instance is connected to the University Hospitals of Leicester Patient Master Index (PMI) ensuring that all recruits to the BRICCS project have accurate matching demographics including hospital identifiers to facilitate acquisition of clinica data from other systems in use in the hospital.

Data collected by OBiBa Onyx in a BRISSKit installation can be encrypted and exported either to an external data repository or to a BRISSKit component such as i2b2 for data warehousing/querying.

A full range of answer types including text, single selection, multiple selection and dropdown lists can be configured. Individual questions, and even whole stages, can be dependent on previous answers, and full branching logic is possible. Extensive validation rules are available.

Multiple administrative data elements (start times, end times, etc.) are captured automatically, and the collection of biosamples can be monitored via text entry or the scanning of barcodes.

Questionnaire stages can be produced in multiple languages, can display images or play audio as part of the questionnaire, and with the appropriate hardware can collect audio files as part of the data collection process. Onyx can be adapted for use on touch-screen interfaces for self-administered scenarios, or use visual categories.

caTissue
http://cabig.nci.nih.gov/tools/catissuesuite

caTissue is a tool for biospecimen inventory management, tracking and annotation. It is part of the Cancer Biomedical Informatics Grid (CaBIG) suite of tools. caTissue permits users to manage data on collection, storage, quality and distribution of specimens. Scalable and configurable, caTissue can support multiple sites, contacts, consents and day-to-day admin work required for research studies. It allows you to manage different types of biospecimens (tissue, biofluids, nucleic acid), and provides search functionality for both the repository staff and research scientists.

caTissue features include:
- Customisable user roles and privileges for multiple, independent repositories within a single installation;

- Collection protocol template creation to expedite data entry;

- Specimen accessioning, recording of processing events, creation and recording of specimen derivatives and aliquots;

- CAP checklist-based forms for pathology annotation of major organ systems;

- caTIES-like pathology report text annotation;

- Advanced Query 'Wizard', allowing for the creation of customised searches

- Ability to bulk load data.

BRISSKit deployments of caTissue can create entries automatically for research subjects recruited and monitored using the CiviCRM component. Barcode readers can be used to enter biospecimen data into caTissue. In addition, BRISSKit uses the extensive caTissue Application Programming Interface (API) to submit data to the caTissue database using a lightweight data-entry program written specifically for the BRICCS project. This speeds up data entry significantly in projects where large numbers of study subjects have multiple samples collected following an identical protocol.

caTissue is particularly well-suited to use across institutions: where multiple storage sites need to be defined, and where moving individual samples (or entire storage containers) needs to be recorded. A full audit trail of all events e.g. moving samples, thawing, freezing, is stored within the caTissue database and visible to system administrators.

i2b2

i2b2
www.i2b2.org

Informatics for Integrating Biology and the Bedside (i2b2) is a scalable informatics framework developed under an open source licence that uses existing clinical data for discovery research, combined with genomic data to facilitate the design of targeted therapies.

i2b2 is a US 'National Center for Biomedical Computing' funded as a cooperative agreement with the national institutes of health and based at Partners HealthCare Systems, Boston, Massachusetts. The i2b2 centre is developing a scalable computational framework to address the bottleneck limiting the translation of genomic findings and hypotheses in model systems relevant to human health.

The platform currently enjoys wide international adoption by the CTSA network, academic health centres, and industry.

i2b2 is funded as a cooperative agreement with the national institutes of health.

The i2b2 architecture is flexible and scalable, allowing for the integration of data from multiple sources into cohort discovery and other data querying functions. A typical BRISSKit use of i2b2 would involve linking questionnaire data generated from an OBiBa Onyx data collection tool with specimen data generated by caTissue. In NIHR Leicester Cardiovascular BRU this data is supplemented by pathology results obtained from the iLab system, and audit data held in the hospital's data warehouse.

Multiple data sources can be loaded against common or individual ontologies within i2b2, and the BRISSKit team can advise further on customisations for specific data sources.

i2b2 is also extensible, with plugins available to manage genomic data within the i2b2 datamart, display health information using the SMART healthcare toolset, or share aggregate research data using the SHRINE data sharing layer. All these enhancements can be implemented optionally onto a BRISSKit deployment.

BRISSKit: A cloud-enabled Translational Research Platform

The Biomedical Research Infrastructure Software Service Kit

The Problem

Translational research bridging the divide between healthcare and biomedical research requires extensive use of computer software to collate, record, integrate and analyse data.

The Solution

BRISSKit will support the management and integration of tissue samples with clinical data and electronic patient records, having appropriate information governance and anonymisation where required. It provides secure, reliable and affordable open source database hosting and cohort selection solutions in-house or as a cloud-based service e.g. over the UK JANET academic network or via secure NHS provider for sensitive data.

Further details available via
www.brisskit.le.ac.uk
or the Project Lead
Dr Jonathan Tedds, brisskit@le.ac.uk

Example BRISSkit Implementation

NHS Host Institution

Patient Master Index

Clinical data

Imaging and other systems

NHS staff and researchers

Relationship management
CiviCRM

Sample tracking
caTissue

Electronic data capture
OBiBa Onyx

Other components
(e.g. REDcap)

Access Control

Identifiable data

Data warehousing
i2b2

Anonymisation

Anonymised data

Researchers

National or international collaboration

Other local researchers

Biomedical Research Group

Access Control

Data querying
i2b2

University of Leicester

JISC

NHS National Institute for Health Research