BRISSkit: Biomedical Research Infrastructure Software Service kit

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

Dr Jonathan Tedds  jat26@le.ac.uk

• UK HEFCE/JISC University Modernisation Fund 2011/12
• Shared service in HE cloud for HE
  • Eduserv cloud platform
  • Brokered by UK HE JANET academic network
UK Research Data Service (UKRDS) pathfinder

- Key finding of HEFCE funded study 2008/9 at Leicester, Oxford, Bristol, Leeds:
  - research facilitators = missing link

- initial funding from Oct 09 + need institutional level support

- Coordinate research life cycle, expertise & training across key institutional stakeholders
  - Researchers in Colleges, depts, groups, projects
  - IT Services – embryonic research computing support – HPC, storage....
  - Research Support Office – funding, legal requirements...
  - Library - repository archive, ethical & training expertise...

- Coordinate to national expertise including
  - JISC Managing Research Data program #jiscmrd
  - Digital Curation Centre – research data management planning tools
  - RCUK & EU FP6,7 e-science program, RIN, OMII, UCISA.....
HALOGEN (History, Archaeology, Linguistics, Onomastics, GENetics): Throwing light on the past through cross-disciplinary databasing

http://halogen.le.ac.uk/index.php

- Portable Antiquities Scheme (British Museum)
- Key to English Place-names (Nottingham)
- Surnames
- Genetics
- IT hosting and GIS
- Best practice:
  #JISCNRD, UKRDS, DCC, international
Suggested institutional timeline

Near term (up to a year)
- Get key players from across institution together: Identify objectives for Research Data Management (RDM)
- Raise awareness and make use of existing general training materials e.g. DCC101
- Identify and benchmark existing central provision for RDM, including research information systems, grant costing & staff support services
- Identify, disseminate and develop current exemplars to conduct basic audit of RDM. Seek quick wins with high profile academic champions
- Draw on external expertise, build links with potential partners

Medium term (1-3 years)
- Establish a formal University wide Committee to develop & ratify policies on RDM; disseminate benefits & cost savings
- Design and pilot institution-wide and discipline-specific training drawing upon local support and guidance
- Appraise RDM provision & conduct gap analysis to develop business case. Benchmark levels of maturity of RDM infrastructure & support
- Develop portfolio of projects to reuse infrastructure, build RDM capacity in further key groups
- Formalise links to external partners & stakeholders where necessary

Long term (>3 years)
- RDM is embedded across faculties & central services, subject to periodic review, ensuring information flow and support is co-ordinated
- Roll out institution-wide and discipline-specific training to all new staff & students. Periodic review and update of materials
- RDM service in place and involving all stakeholders. Periodic benchmarking and target setting
- RDM and Policies developed at faculty level; Key academic groups are engaged; central services in widespread use across the institution
- RDM effectively integrated with shared UK and international services; quality assurance in place

From Whyte & Tedds (2011), DCC Briefing
http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm
**BRISSkit Generic Use Cases 1**

- **Researcher**
  - Designs study
  - Identifies potential patient profiles and authorises patient recruitment process
  - Authorises nurse and research informatics consent, tissue and data gathering activity
  - Analyses collected data and conducts study
  - Publishes study results
  - Transfers ownership of data BRU on completion of study

- **Research Informatics Lead**
  - Designs questionnaire and hospital dataset
  - Maps required ontology to dataflows
  - Authorises hospital informatics activities
  - Ensures readiness of questionnaires, samples and data gathering interfaces
  - Monitors patient consent to enable start/stop clinical data gathering and use of samples
BRISSkit Generic Use Cases 2

- **Patient**
  - Consents to participate in research/study
  - Provides data through questionnaires
  - Provides tissue samples
  - Provides data through clinical systems

- **Nurse**
  - Gains patient consent to participate
  - Supports patient completion of questionnaires
  - Takes blood and tissue samples from patient
  - Orders other clinical tests on patient

- **Clinical Informatics Lead**
  - Monitors existence of patient consent
  - Ensures readiness of clinical systems to share data
  - Initiates, monitors and closes sharing of patient data into research system for consenting patients
BRISSkit: Biomedical Research Infrastructure Software Service kit

- Electronic Data Capture / questionnaire – OBiBa Onyx / REDcap
- Specimen Inventory / sample tracking – caTissue
- Cohort selection and data querying - i2b2
BRISSkit: Biomedical Research Infrastructure Software Service kit

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

- Prototype open source IT infrastructure for biomedical research Informatics
- Web services – platform independent
- Bridge healthcare & research domains
  - *i4health international & ITFoM EU initiatives*
- Host, implement, deploy biomedical research DBs
  - *Remove duplication* of highly skilled development effort via
  - Design from scratch
  - Buy & administer
Electronic Data Capture

**Onyx** - from the **OBiBa Open BioBank suite of software tools** - manages baseline interviews.

Records participant consent, questionnaire data & primary specimen IDs.

Web-based, secure data entry by research staff. Used for all recruits to BRICCS prototype - mobile computing on wards and outpatient clinic.

Ontario/Montreal/Perth (Aus)/East Coast US

Allows for integration with physical measurement devices, manages appointments, participant look-up to registry services or walk-in volunteers.
**Electronic Data Capture**

**REDCap** allows for common data elements to be re-used in building and managing online surveys and databases.

Create and design projects using your web browser using the Online Designer; and/or by constructing a 'data dictionary' template file in Microsoft Excel.

Both surveys and databases (or a mixture of the two) can be built.

**REDCap** provides audit trails for tracking data manipulation and user activity, as well as automated export procedures for seamless data downloads to Excel, PDF, and common statistical packages (SPSS, SAS, Stata, R).
Specimen Management

cаТissue is caBIG's biorepository tool for biospecimen inventory management, tracking, and annotation. This tool permits users to enter and retrieve data concerning the collection, storage, quality assurance, and distribution of biospecimens.

cаТissue Suite is scalable and configurable for deployment across biospecimen resources of varying size and function, and that manage multiple types of biospecimens (tissue, biofluids, nucleic acid).

Recording of biospecimen shipping and tracking events across repositories in a single caTissue installation.
Translational Data Management

I2b2 is a comprehensive software and methodological framework to enable clinical researchers to accelerate the translation of genomic and “traditional” clinical findings into novel diagnostics, prognostics, and therapeutics.

Data from disparate sources can be loaded into i2b2 under a sophisticated mechanism of ontology building, allowing for complex searching and cross-referencing of data, visualisations and export.

Plugins allow for natural language processing and data visualisation.
Central Participant Registration

C3PR - caBIG Central Clinical Participant Registry - enables efficient and streamlined registration of participants into studies. It can be used by an individual site, or by a multi-institutional organization with many, graphically dispersed sites.

C3PR helps to organize, and standardize a template for patient registration, informed consent, inclusion/exclusion criteria, stratification categories, and treatment arms.

The system tracks screening failures and enrollment statistics, and can be configured to alert study personnel when accrual thresholds are met.
BRISSkit Virtual Machines

Master Patient Index
Use SNO & BPT

C3PR uses new Contact no, SNO, BPT

ONYX
Uses BPT

CLINICAL SYSTEMS

XDS
Uses BPT

REDCAP
Use SNO

ONYX
Use BPT

caTISSUE
Uses BPT

i2b2 N3 SNO

i2b2 Eduserv
Uses BPT

Single Public Research Facility
With anonymised data and samples

Single Instance for Biomedical Research Unit Group

BRISSkit Admin Portal

Web services + XML documents
Leicester N3
Virtual Machines
Eduserv
Virtual Machines
N3 Fixed Network Apps

Multiple Biomedical Research Centres
End-to-end informatics support within Leicester

Open source infrastructure

Cloud Services offered to other Biomedical Research Units (NHS-HE):

- Client Analytics
- Client search, selection and recruitment
- Client Data Integration
- Client Engagement
- Client Asset Management
Service Options

Client Data Analytics (i2b2)

Client Search, Selection and Registration (C3PR+)

Client Data Integration and Audit (XDS)

Client Engagement (Onyx)

Client Asset Management (caTissue)
BRISSkit Sustainability - OS engagement

• OS community engagement
  • standards compliance
  • service vision

• Cross-enterprise Service Architecture:
  • how to join & use service for new groups & partners
  • Definition of service vision, organisational & service components
  • all relevant standards & tools which Brisskit partners will be expected to use and comply with

• OS Community Engagement Charter
  • defining engagement with existing & new OS communities
  • including adoption & code commitments
BRISSkit - meeting a challenge?

Is there value in delivering cloud-based applications as a service to researchers?

- Ease of deployment, administration and archiving
- Cost of computing can be scaled to projects
- Model lends itself to distributed projects and teams
- Infrastructure can be geared to integration
- Enhanced basis for data sharing and federation
- Need to ensure sustainability of the service
- Confidentiality and governance

National Workshop - Thurs 19th January, Leicester

Jonathan Tedds  jat26@le.ac.uk  @jtedds  http://www.le.ac.uk/BRISSkit
DIRECT 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
- 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: EU FP7 Mintweld (industrial engineering) & BRICCS National Health Service/University Trust data sharing.

*Increasing skills base of researchers/students/staff*
ORGANISATIONAL CHALLENGES AND SOLUTIONS

Cultural differences

- Recognise different cultures and mind sets
  - research community and IT specialists in central services
  - different professional language, expectations and working practises
  - management of a research project usually requires a different, iterative methodology than an IT infrastructure project having a more clearly pre determined end point

Research Liaison Role

- An IT specialist with strong research background
- enables effective ways of liaising with research community
- bridging gaps in understanding

Leveraging expertise within and external to the organisation

- coordinate ‘specialists’

See Research Fortnight blog piece Feb 2011
### BRISSkit: Biomedical Research Infrastructure Software Service Kit

#### Theme

<table>
<thead>
<tr>
<th>Core Components</th>
<th>Lead</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Service Design</strong></td>
<td></td>
</tr>
<tr>
<td>Detailed project plan (1.1)</td>
<td></td>
</tr>
<tr>
<td>Engagement with key stakeholders (UoL / UHL / NHS / JISC / JANET / Eduserv / International)</td>
<td></td>
</tr>
<tr>
<td>Revised project plan (1.2)</td>
<td></td>
</tr>
<tr>
<td>Provision of Workshops (9.1, 9.2-9.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Technical development</strong></td>
<td></td>
</tr>
<tr>
<td>Generic deployment infrastructure VM Test facilities (2.2)</td>
<td></td>
</tr>
<tr>
<td>caTissue API client (4.2)</td>
<td></td>
</tr>
<tr>
<td>Data generation tools (5.4)</td>
<td></td>
</tr>
<tr>
<td>Data warehouse application (7.3)</td>
<td></td>
</tr>
<tr>
<td>UK customisations (7.1)</td>
<td></td>
</tr>
<tr>
<td>Integration of caTissue / i2b2 (8.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Return on Investment</strong></td>
<td></td>
</tr>
<tr>
<td>Baseline – NIHR BRU / BRC</td>
<td></td>
</tr>
<tr>
<td>Baseline – non NIHR Biomedical Research Groups</td>
<td></td>
</tr>
<tr>
<td>Model cost savings and develop Business Plan</td>
<td></td>
</tr>
</tbody>
</table>

#### Lead

- Jonathan Tedds (UoL) / Malcolm Newbury (GuildFoss) / Chris Greengrass (UHL BRICCS) / Nilesh Samani (UoL Cardiovascular) / Tony Brookes (UoL Genetics) / Paul Burton (UoL Health Sciences) / Peter Knight (NIHR) / Kevin Harris (UHLT) / Kevin Schurer (UoL PVC Research) / JISC / JANET / Eduserv / HEFCE
- Nick Holden (UHL) / Jeff Lusted (UoL) / Debadutt Goswami & team (UoL)
- Charles Beagrie Consultants / JISC / JANET / Eduserv / HEFCE / UoL Business Development