Research Data Management at Leicester

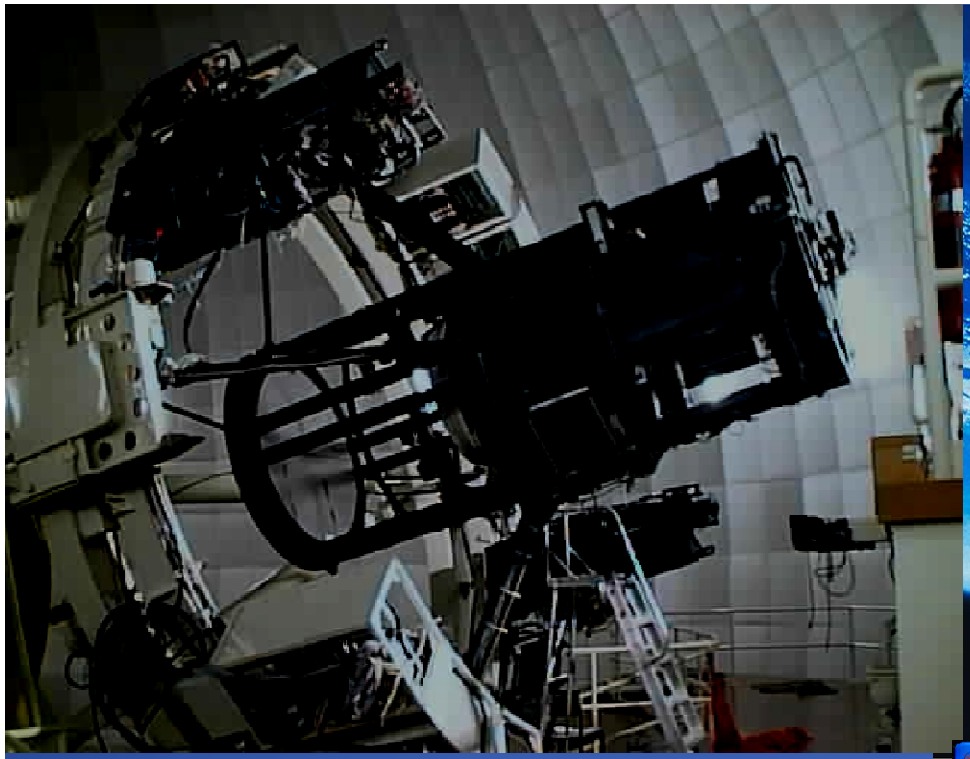
Dr Jonathan Tedds

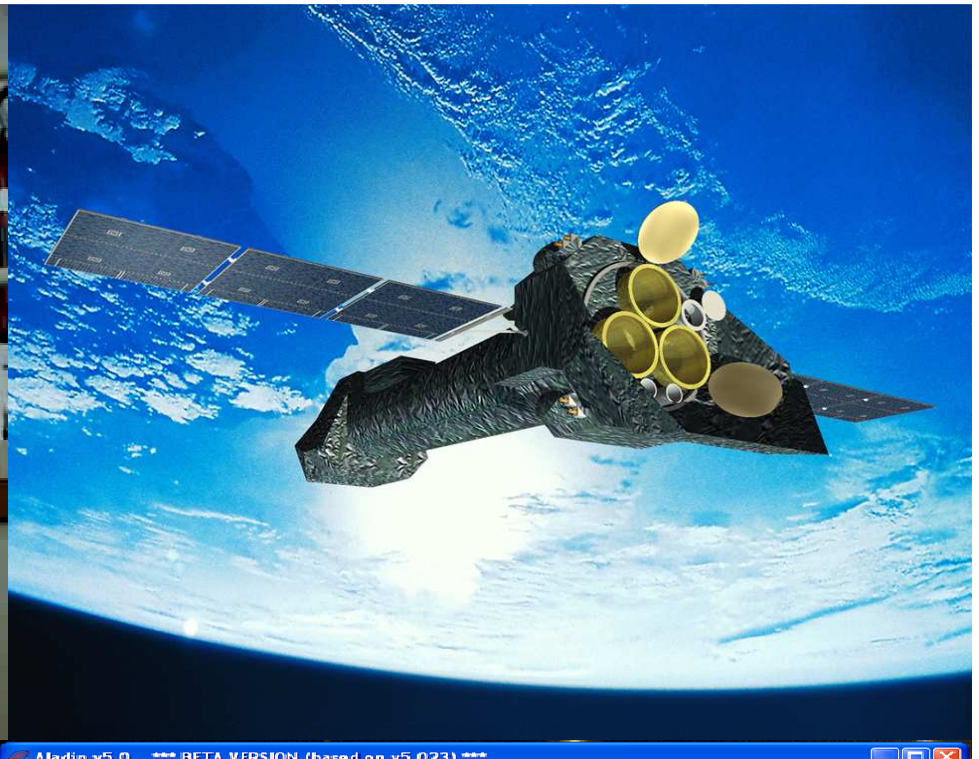
University of Leicester
Senior Research Liaison Manager
IT Services
Research Fellow
Dept of Physics & Astronomy

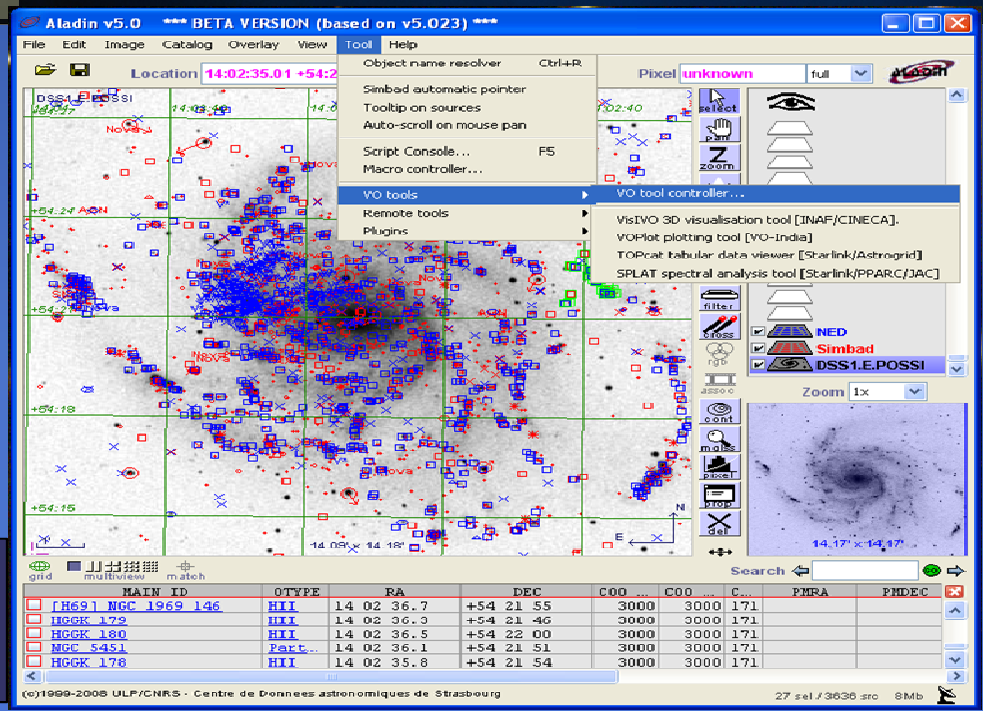
twitter @jtedds

jat26@le.ac.uk

University of Leicester



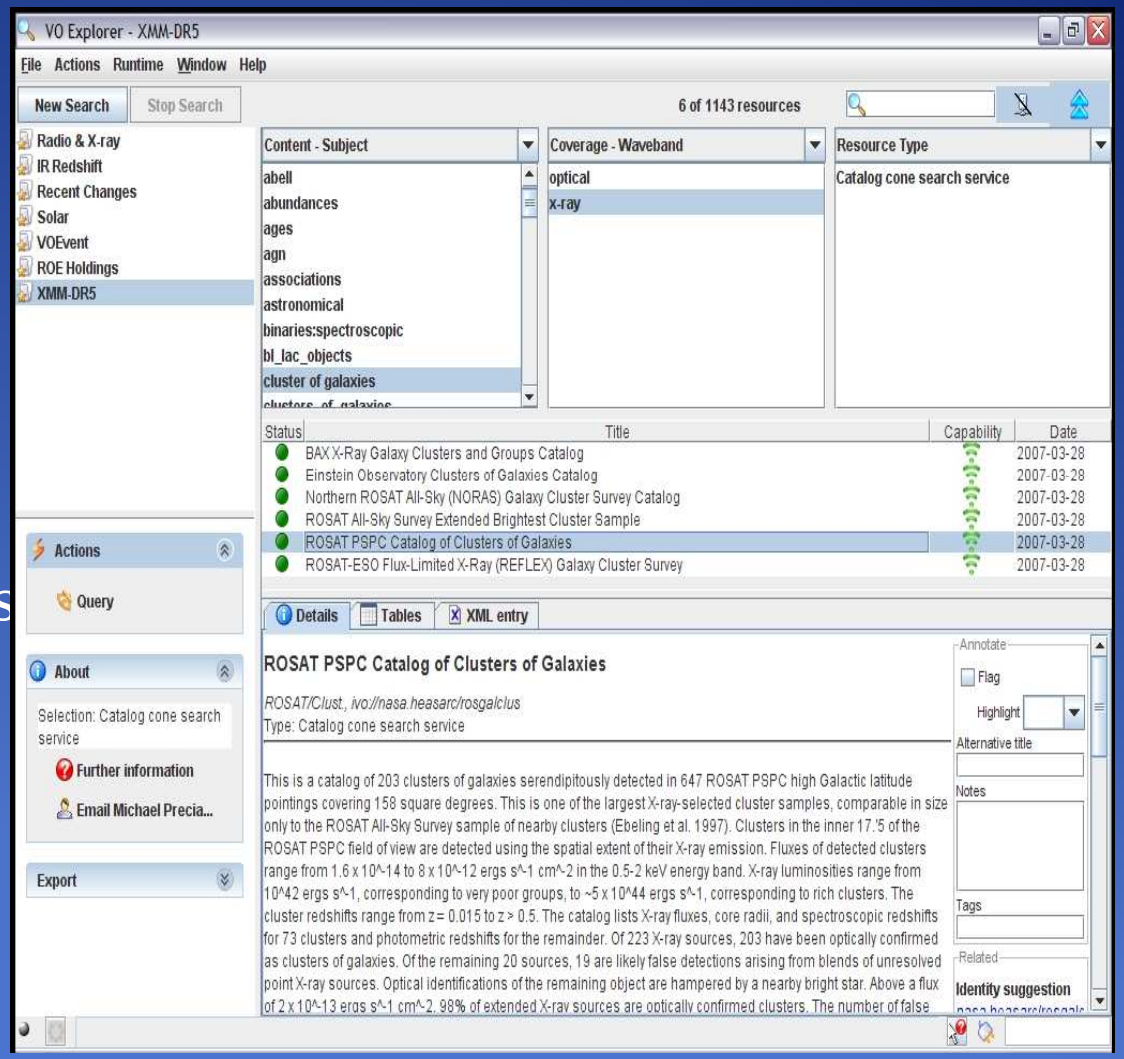




http://www.astrogrid.org
(April 2008 1st public release)

http://www.astrogrid.org

- AstroGrid front end, part of VOdesktop
- *Resource-centric*
- Select a search-space
- Search for resources
- Filter these resources
- View selected resources
- Use the selection
  - Invoke it
  - Save/Bookmark/Tag it
  - Export it



University of Leicester

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

University of Leicester

wellcome trust

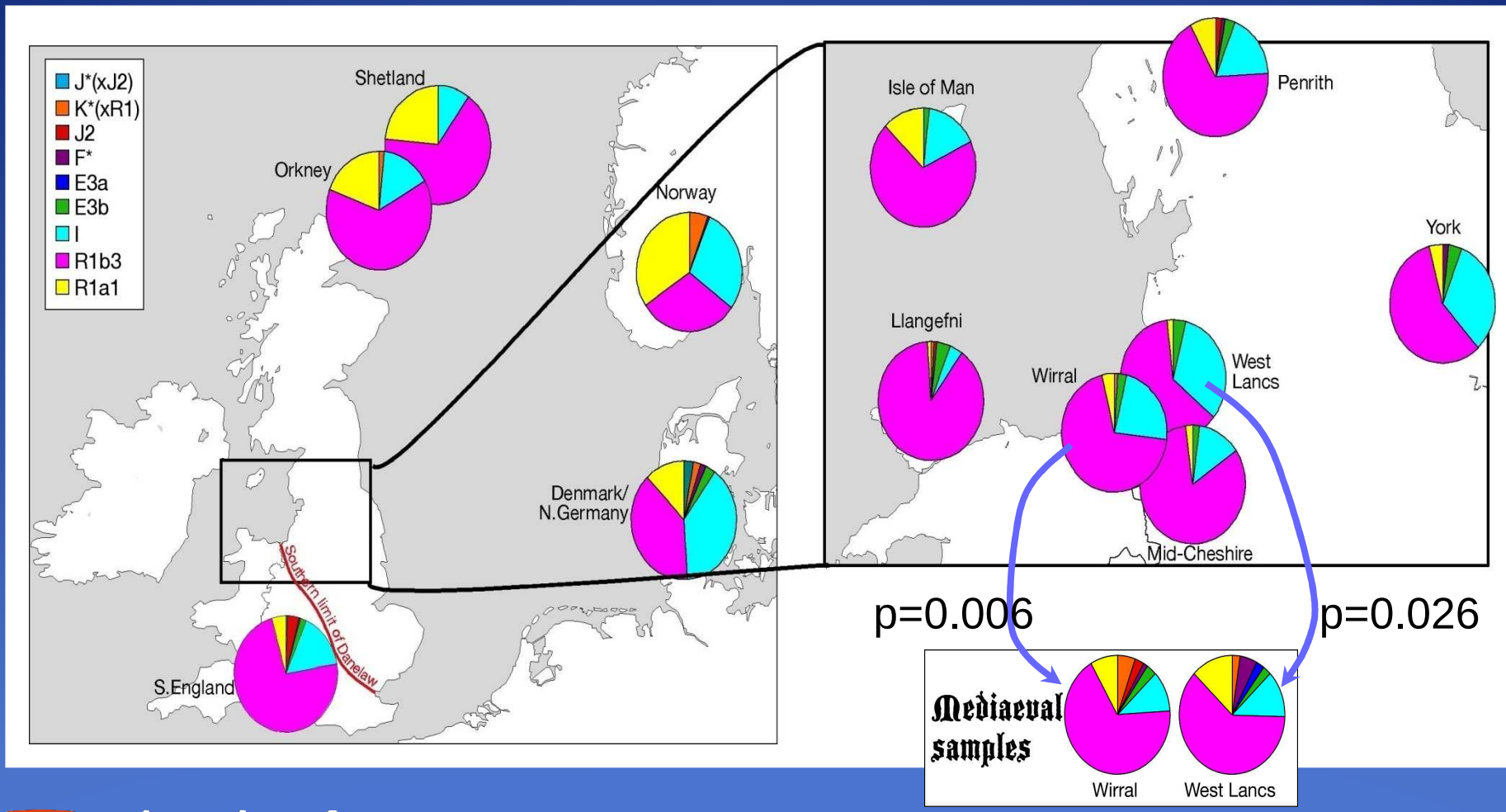
- Portable Antiquities Scheme (British Museum)
- Place-names (Nottingham)
- Surnames
- Genetics
- IT hosting and GIS
- Best practice: #JISCMRD, UKRDS, DCC, international

http://halogen.le.ac.uk

JISC

Bowden *et al*., 2008

Genetic variation (Y data)



- Population differentiation test

University of Leicester

Halogen as template for research data management #jiscmrd

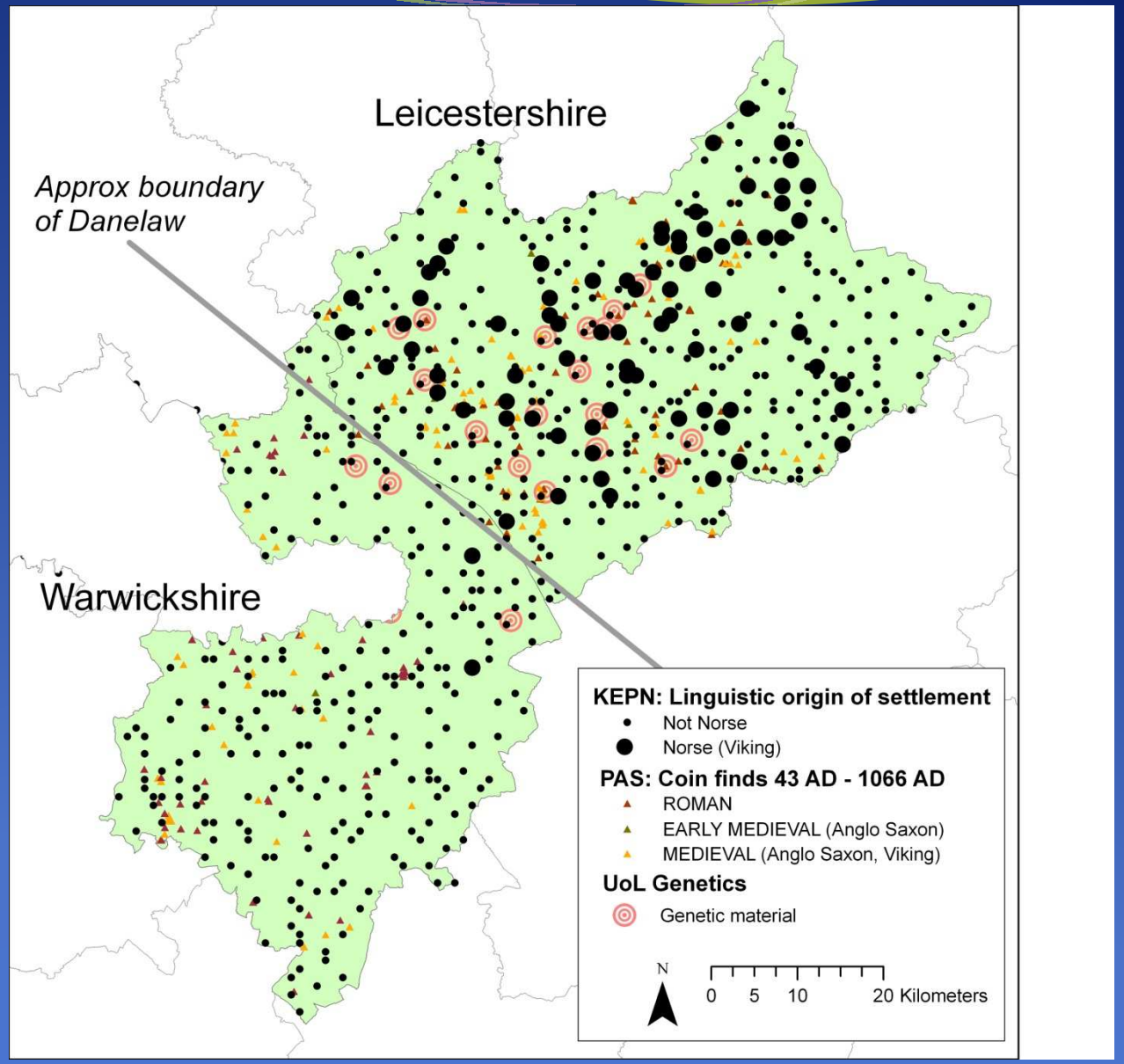
- Requirements Analysis – must be iterative!
- *Data Management Plan – use DMPonline (DCC)*
- Scalable research data management infrastructure
  - pilot phase to nationally available resource
  - LAMP stack IT infrastructure: host research database – work with JISC/DCC
- A model for the long term delivery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

University of Leicester

Halogen output combines input datasets via Geographical Information Systems (GIS):

- linguistics
- archaeology
- genetics

=> #JISCgeo programme



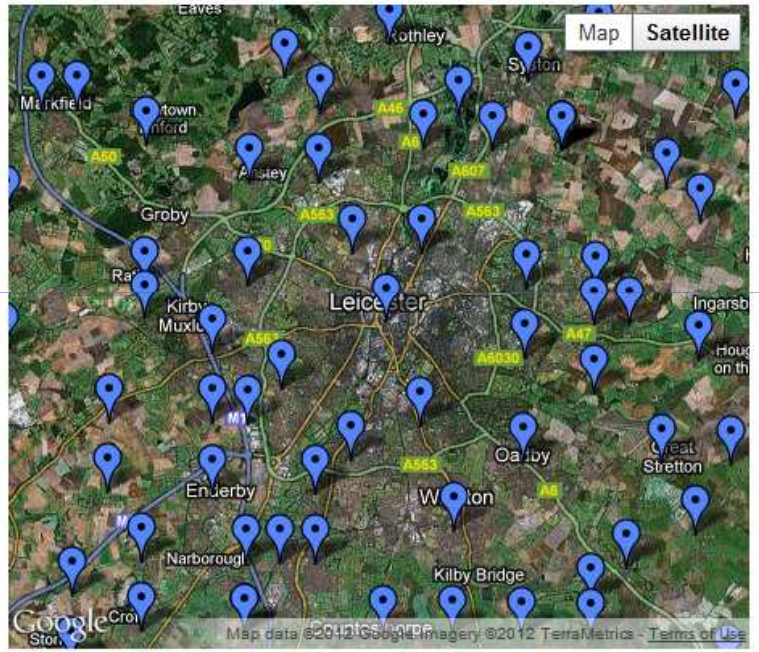
University of Leicester

University of Leicester

HALOGEN

HALOGEN Home | Query | Guide | FAQ | Partners | License | Contact us

Welcome to the HALOGEN geospatial search facility



Key to English place name data around Leicester.

We host geospatial data covering the etymology of English place names, the 1881 census, genetics data and the portable antiquities data. A more in depth summary of each data set is available in the user guide

The map on the left shows an example of the data available near you. You can search specific data sets by following the links below.

Do a English place-name etymology query.
Do a genetics query.
Do a census surname query.
Do a census parish query.
Do a portable antiquities query.

Further information about the HALOGEN project can be found at www.le.ac.uk/halogen.

JISC | The University of Nottingham | The British Museum

Legal

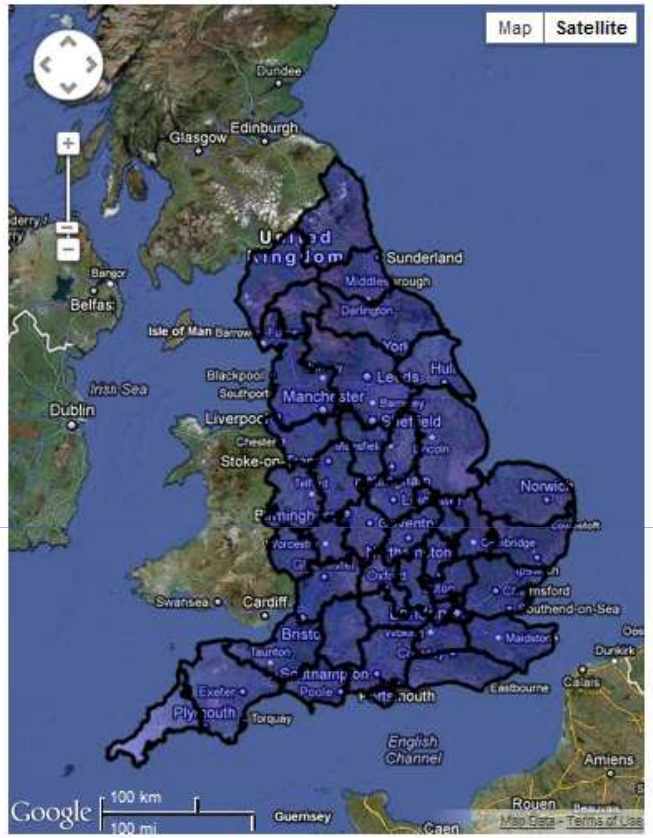
University of Leicester

HALOGEN

HALOGEN Home | Query | Guide | FAQ | Partners | License | Contact us

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

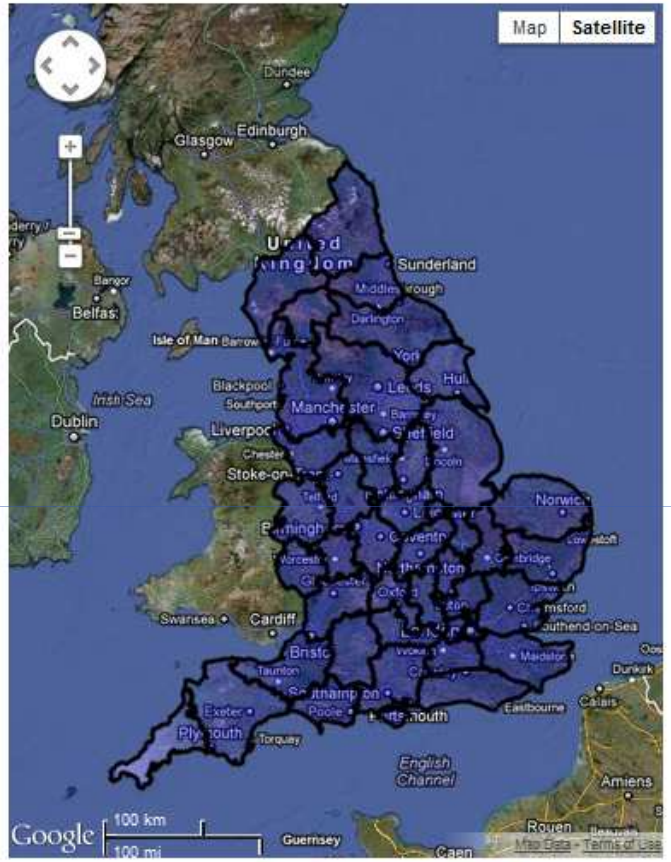
University of Leicester

JISC | The University of Nottingham | THE BRITISH MUSEUM

HALOGEN

HALOGEN Home | Query | Guide | FAQ | Partners | License | Contact us

Portable Antiquities Scheme



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 period:

ALL PERIODS

Bronze age
Byzantine
Early Medieval
Foreign
Greek and Roman pro
Iron age

Submit Reset

JISC The University of Nottingham THE BRITISH MUSEUM

University of Leicester

CHALLENGES

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

University of Leicester

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*

University of Leicester

Reward = Leverhulme Trust funding £1.3m!

Indirect Benefits (Costs Avoided)

- *No re-creation of data*
  - Researchers avoid valuable time needed to transcribe external data sources
- *Inter disciplinary research platform available centrally for reuse as a service*
- *Lower future preservation costs*
  - Reusable Service Level Agreements in place
  - Not dependent on individuals alone
- *Re-purposing data, methodologies for new audiences*
  - Internal & national research resources can become nationally reusable
  - e.g. Geneticists learn better spatial correlation analysis techniques
- *Protecting returns on earlier investments*
  - research funders: Wellcome Trust, Leverhulme Trust, AHRC, British Museum
  - Institutions: Universities of Leicester, Nottingham, UCL

Biomedical Research Infrastructure Software Service: **BRISSkit**

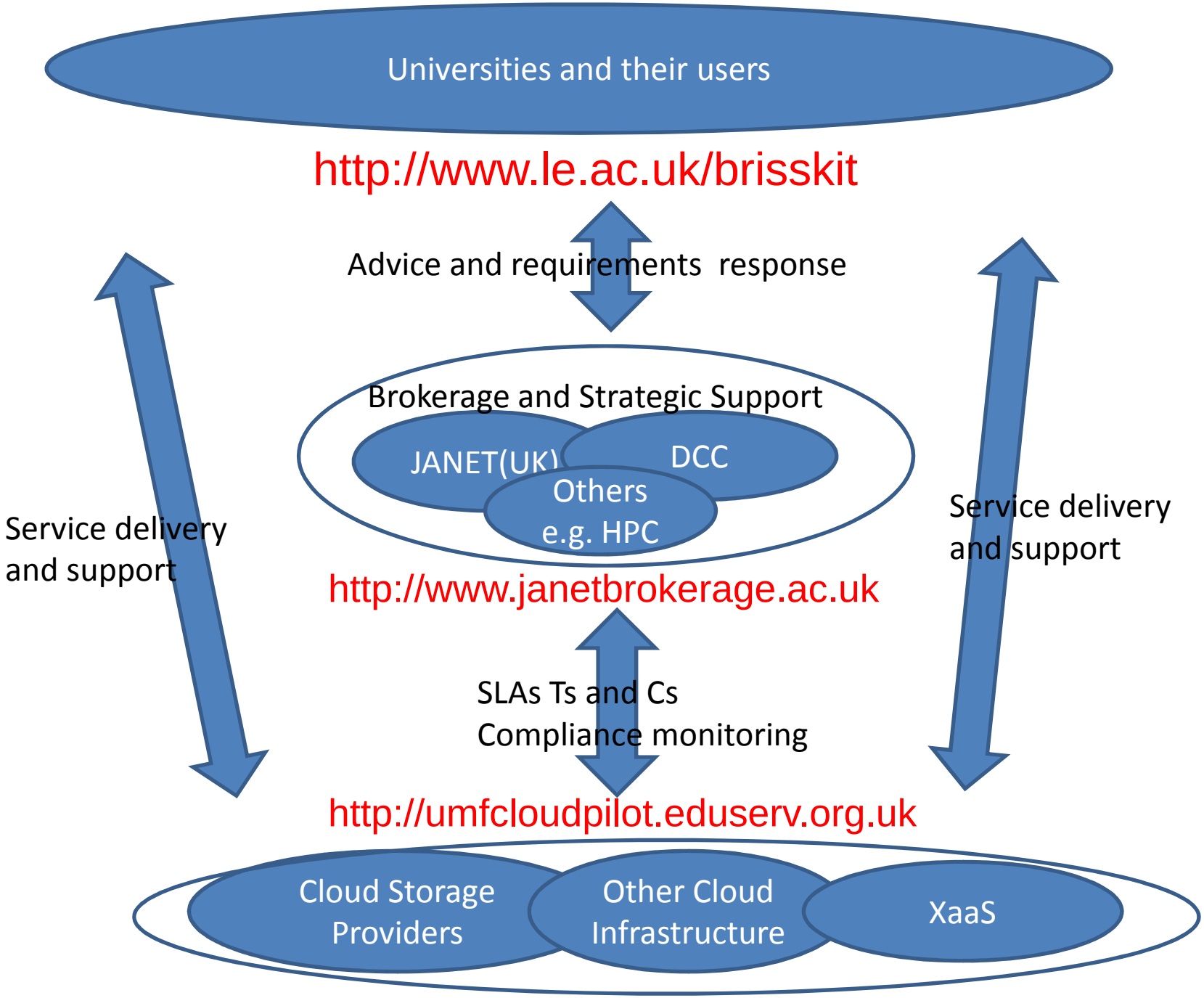
http://www.le.ac.uk/brisskit

- JISC University Modernisation Fund 2011/12
- UoL ITS-RCS, Cardiovascular, Genetics & Health Sciences + UHL Trust
- *develop national shared service in cloud with Eduserv on VMware*
- *brokered by JANET*
- host, implement, deploy biomedical research database applications
  - *Remove duplication of highly skilled development effort via*
    - *Design from scratch*
    - *Buy & administer*
- developed prototype open source IT infrastructure for the Biomedical Research Informatics Centre for Cardiovascular Science in the NIHR Leicester Cardiovascular Biomedical Research Unit
  - *Anonymised versions of patient sensitive databases are hosted for University researchers on LAMP VMware platform implemented by Research Computing Service in IT Services*

University of Leicester

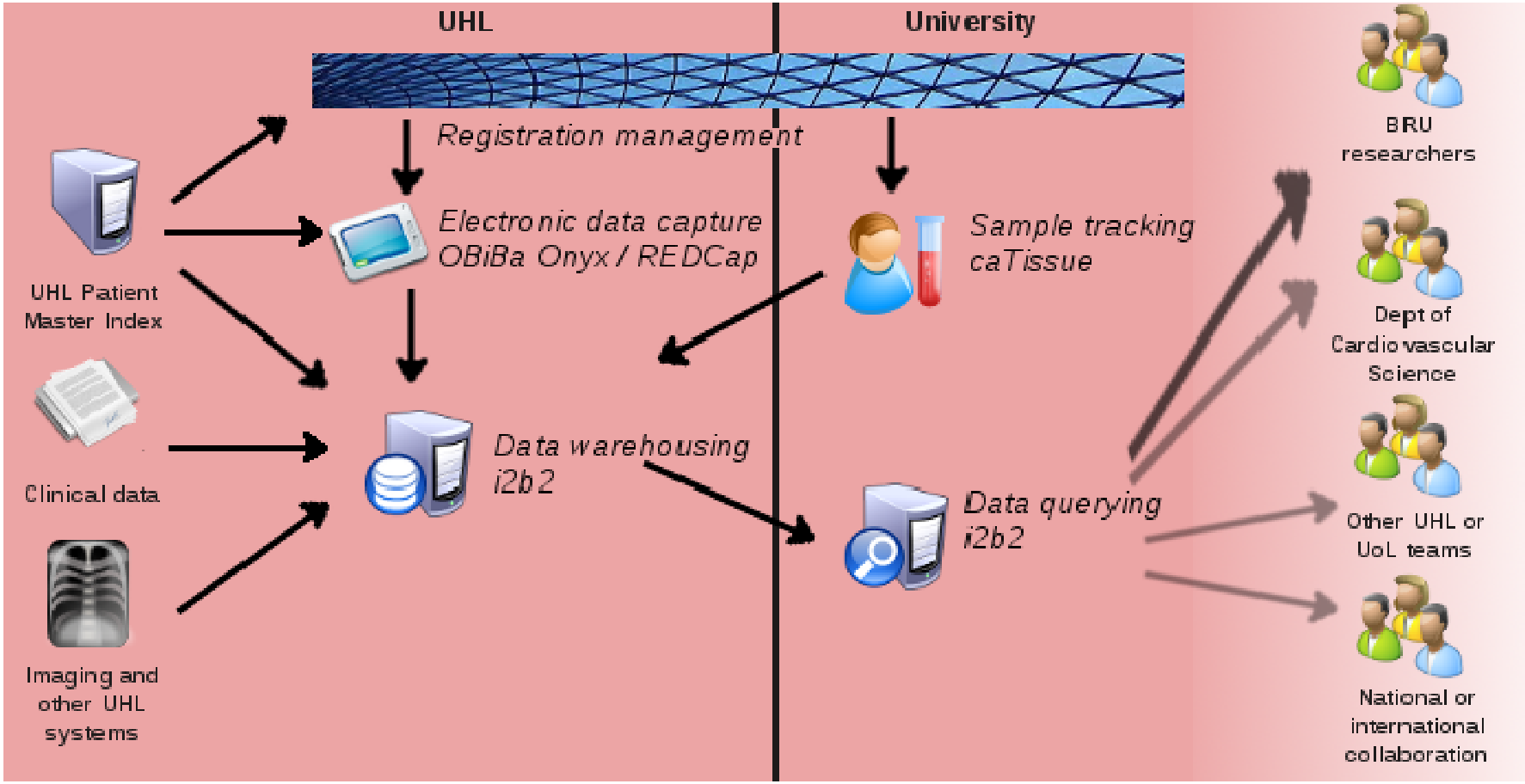
JISC

National Institute for Health Research
Leicester Cardiovascular Biomedical Research Unit


Universities and their users
http://www.le.ac.uk/brisskit
Advice and requirements response
Brokerage and Strategic Support
JANET(UK)
DCC
Others
e.g. HPC
Service delivery
and support
Service delivery
and support
http://www.janetbrokerage.ac.uk
SLAs Ts and Cs
Compliance monitoring
http://umfcloudpilot.eduserv.org.uk
Cloud Storage
Providers
Other Cloud
Infrastructure
XaaS

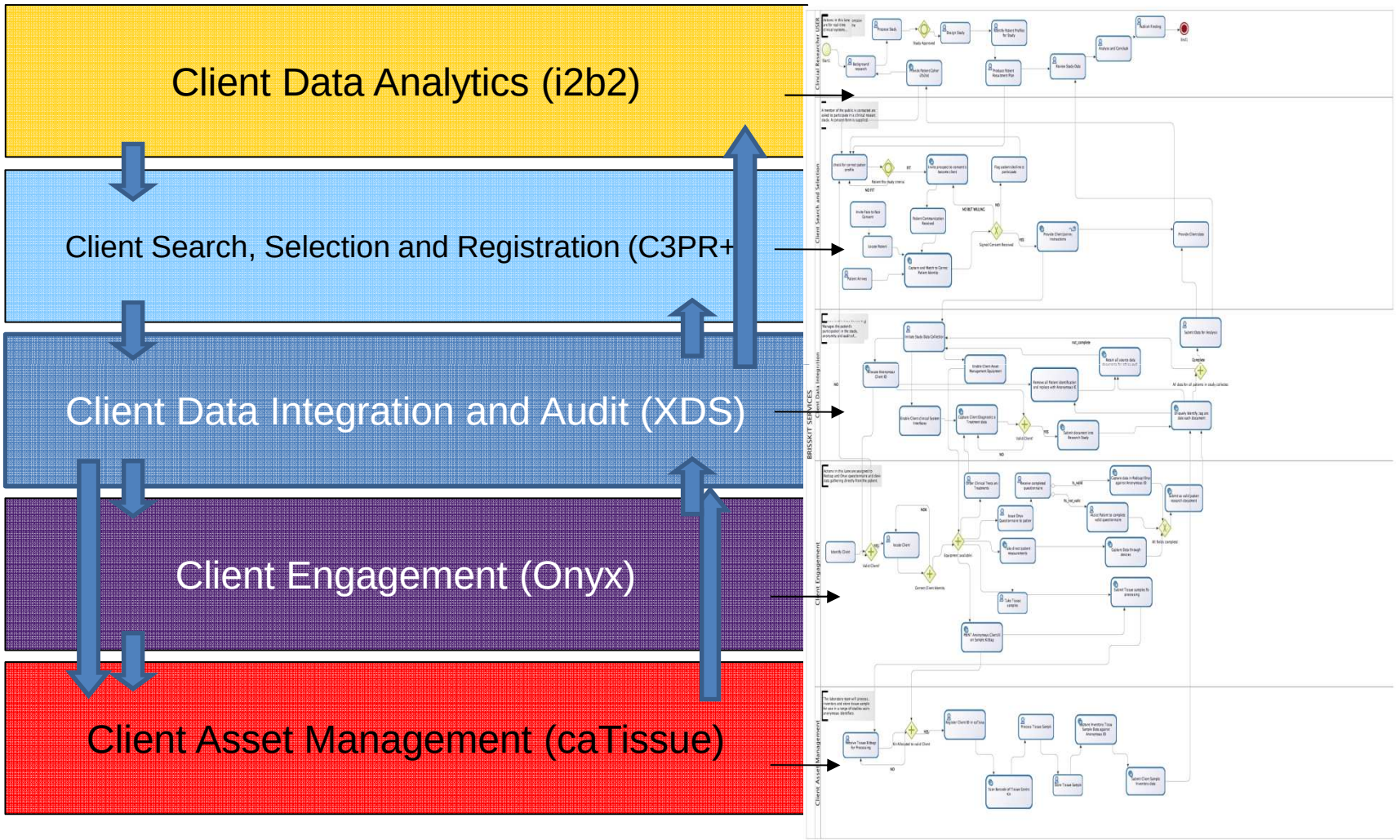
BRISSkit: Biomedical Research Infrastructure Software Service kit

- Electronic Data Capture / questionnaire – *OBiBa Onyx / REDcap*
- Specimen Inventory / sample tracking – *caTissue*
- Cohort selection and data querying - *i2b2*



Service Options

NHS
National Institute for Health Research



PROCESSES

NHS
National Institute for Health Research

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?

National Institute for Health Research

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

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

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*
  - strong research background helps!
  - 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

University of Leicester

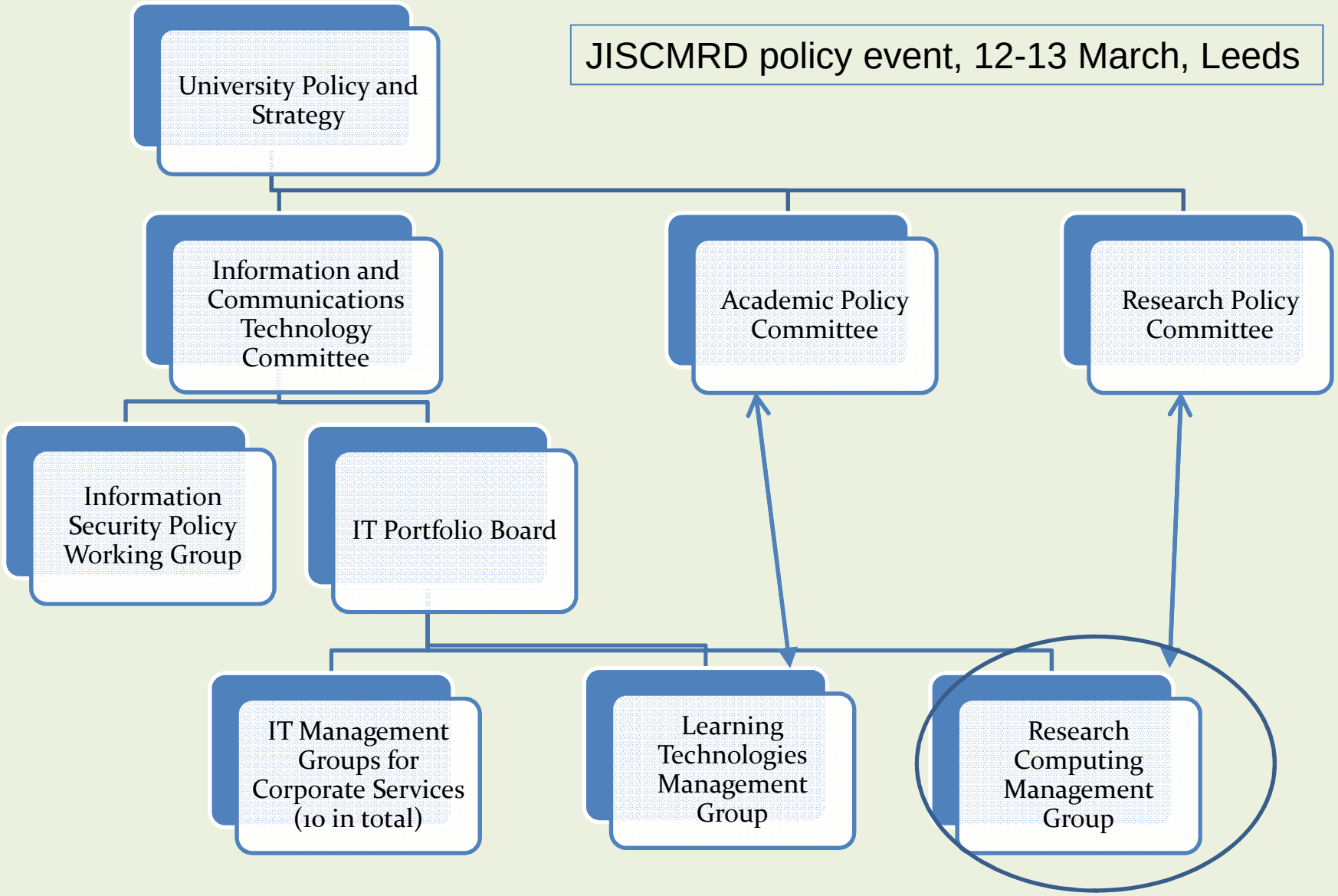
Top Tip: how to get researchers' attention?

- Research grant pre-award costing (LUCRE - SAP based)
  - Dominates researchers' minds!
  - Enable PIs to build grant application using actual costs of staff, overheads, and the right rules for funder
  - Trigger involvement of IT Research Liaison and wider institutional expertise via flags
    - sensitive research data
    - costing/planning support including curation and preservation over research lifecycle
    - Track institution wide needs via IT Service Desk

University of Leicester

Governance

JISCMRD policy event, 12-13 March, Leeds



Suggested institutional timeline

Near term (up to a year)	Medium term (1-3 years)	Long term (>3 years)
Get key players from across institution together: Identify objectives for Research Data Management (RDM)	Establish a formal University wide Committee to develop & ratify policies on RDM; disseminate benefits & cost savings	RDM is embedded across faculties & central services, subject to periodic review, ensuring information flow and support is co-ordinated
Raise awareness and make use of existing general training materials e.g. DCC101	Design and pilot institution-wide and discipline-specific training drawing upon local support and guidance	Roll out institution-wide and discipline-specific training to all new staff & students. Periodic review and update of materials
Identify and benchmark existing central provision for RDM, including research information systems, grant costing & staff support services	Appraise RDM provision & conduct gap analysis to develop business case. Benchmark levels of maturity of RDM infrastructure & support	RDM service in place and involving all stakeholders. Periodic benchmarking and target setting
Identify, disseminate and develop current exemplars to conduct basic audit of RDM. Seek quick wins with high profile academic champions	Develop portfolio of projects to reuse infrastructure, build RDM capacity in further key groups	RDM and Policies developed at faculty level; Key academic groups are engaged; central services in widespread use across the institution
Draw on external expertise, build links with potential partners	Formalise links to external partners & stakeholders where necessary	RDM effectively integrated with shared UK and international services; quality assurance in place

University of Leicester

From Whyte & Tedds (2011), DCC Briefing

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

University *of*
Leicester

Priorities for IT service enhancement and future investment in research computing

- Storage & curation
  - Research data management & planning - including sensitive data
  - LINUX platform, LAMP stack database hosting
  - training
- Performance
  - HPC on demand, network fit for purpose
- Enable Collaboration
  - Internal/external (including non HEIs)
  - Across ITS, RSO, Library for researchers
- Coordinate Expertise
  - Cross disciplinary - Halogen (GIS), BRICCS (astronomy e-science)
  - Grant bidding - IT costing & support
  - Best practice (JISC, DCC, UKDA..international)

University of Leicester

Levels of funding for BRCs and BRUs