The Halogen Project & Roots of the British interdisciplinary research at Leicester

Dr Jonathan Tedds

University of Leicester
Senior Research Liaison Manager
IT Services

twitter @intemple

jat26@le.ac.uk

University *of* Leicester

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, RIN, internatlional

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

JISC

Halogen as template for research data management #jiscmrd

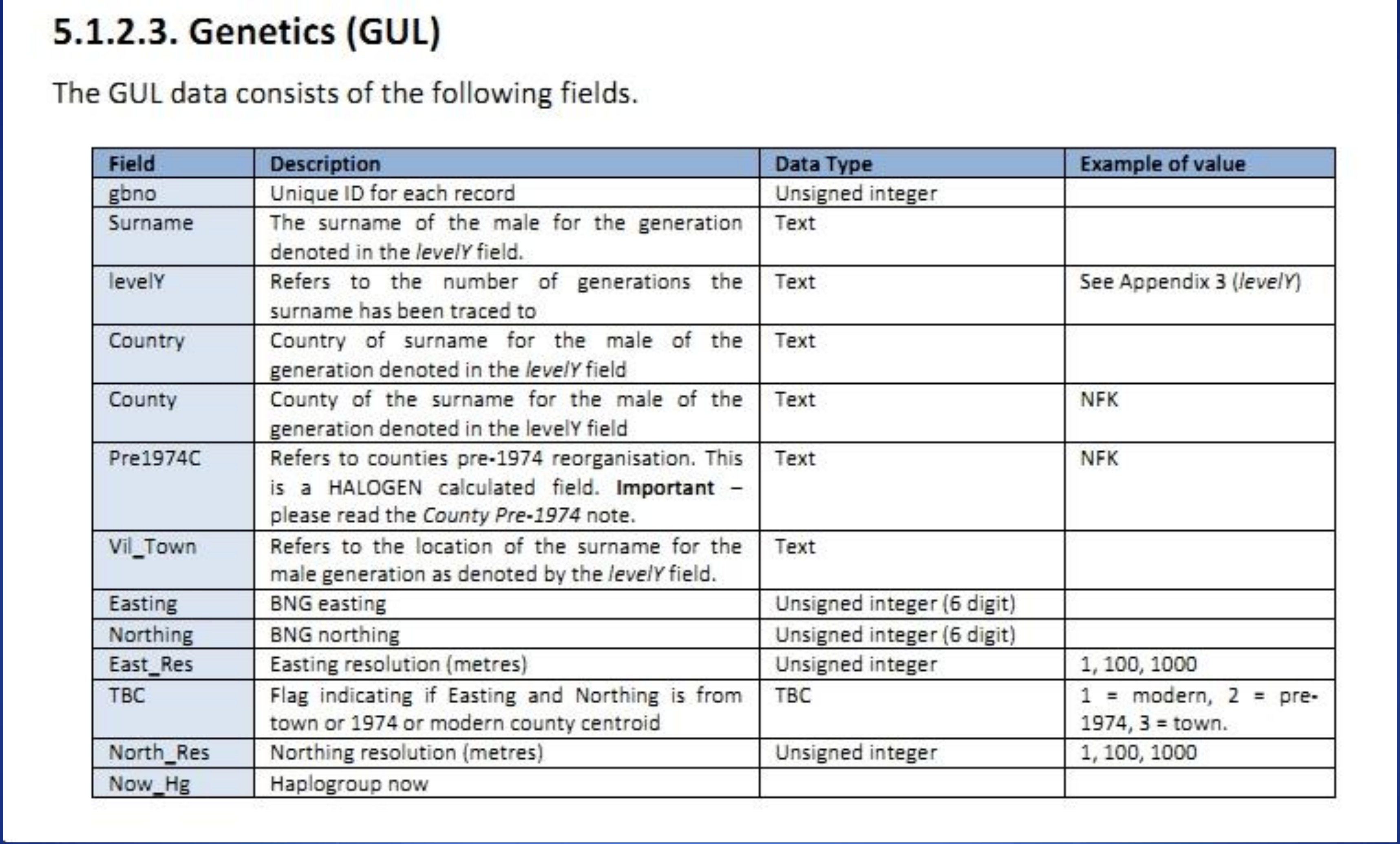
- Requirements Analysis – must be iterative!

- *Data Management Plan – use DMPonline (DCC)*
  - Emphasis on funder rules e.g. Wellcome, PAS, AHRC…
  - Derived from "Data Glossary" document ~44 pages for 3 input sources so far!

- 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

Halogen Data Glossary



5.1.2.3. Genetics (GUL)

The GUL data consists of the following fields.

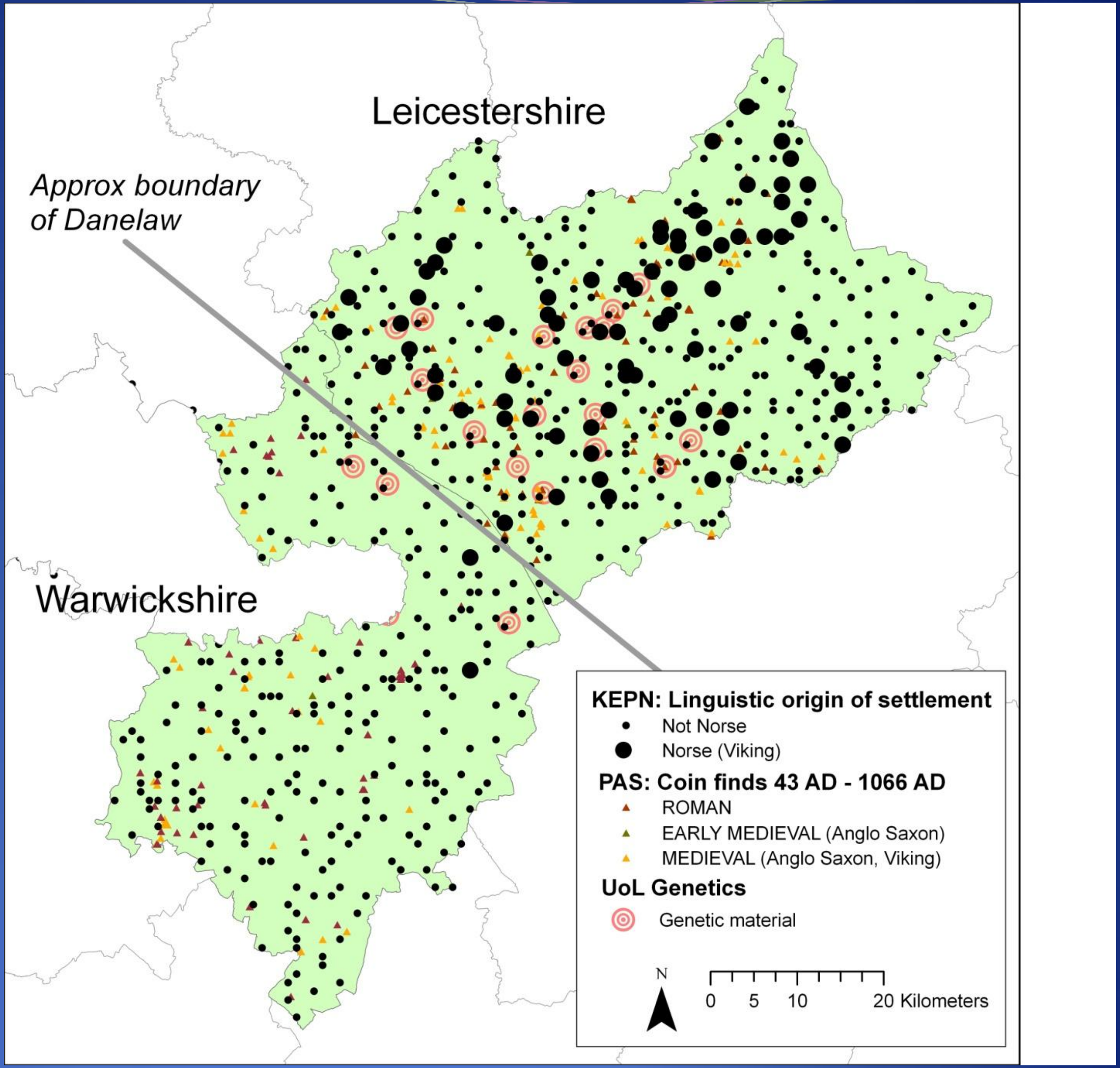
Field	Description	Data Type	Example of value
gbno	Unique ID for each record	Unsigned integer	
Surname	The surname of the male for the generation denoted in the *levelY* field.	Text	
levelY	Refers to the number of generations the surname has been traced to	Text	See Appendix 3 (*levelY*)
Country	Country of surname for the male of the generation denoted in the *levelY* field	Text	
County	County of the surname for the male of the generation denoted in the levelY field	Text	NFK
Pre1974C	Refers to counties pre-1974 reorganisation. This is a HALOGEN calculated field. **Important** – please read the *County Pre-1974* note.	Text	NFK
Vil_Town	Refers to the location of the surname for the male generation as denoted by the *levelY* field.	Text	
Easting	BNG easting	Unsigned integer (6 digit)	
Northing	BNG northing	Unsigned integer (6 digit)	
East_Res	Easting resolution (metres)	Unsigned integer	1, 100, 1000
TBC	Flag indicating if Easting and Northing is from town or 1974 or modern county centroid	TBC	1 = modern, 2 = pre-1974, 3 = town.
North_Res	Northing resolution (metres)	Unsigned integer	1, 100, 1000
Now_Hg	Haplogroup now		

University *of* Leicester

wellcome trust

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

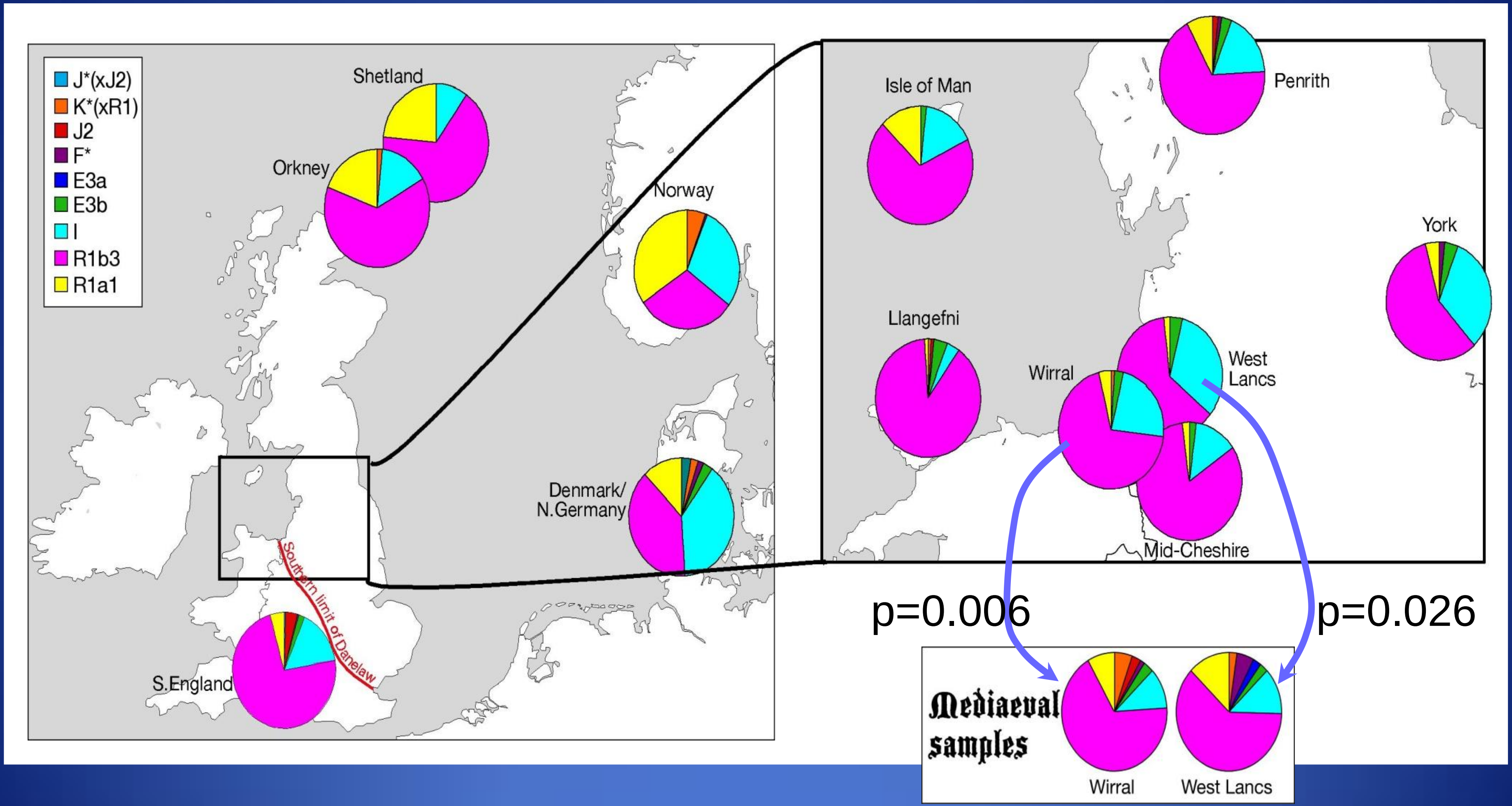
- linguistics
- archaeology
- genetics


Leicestershire
Approx boundary of Danelaw
Warwickshire
KEPN: Linguistic origin of settlement
● Not Norse
● Norse (Viking)
PAS: Coin finds 43 AD - 1066 AD
▲ ROMAN
▲ EARLY MEDIEVAL (Anglo Saxon)
▲ MEDIEVAL (Anglo Saxon, Viking)
UoL Genetics
◎ Genetic material
N
0 5 10 20 Kilometers

University of Leicester

Bowden *et al.*, 2008

Genetic variation (Y data)



- Population differentiation test

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

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!

The Impact of Diasporas *on* The Making of Britain

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

Organisational Challenges and Solutions

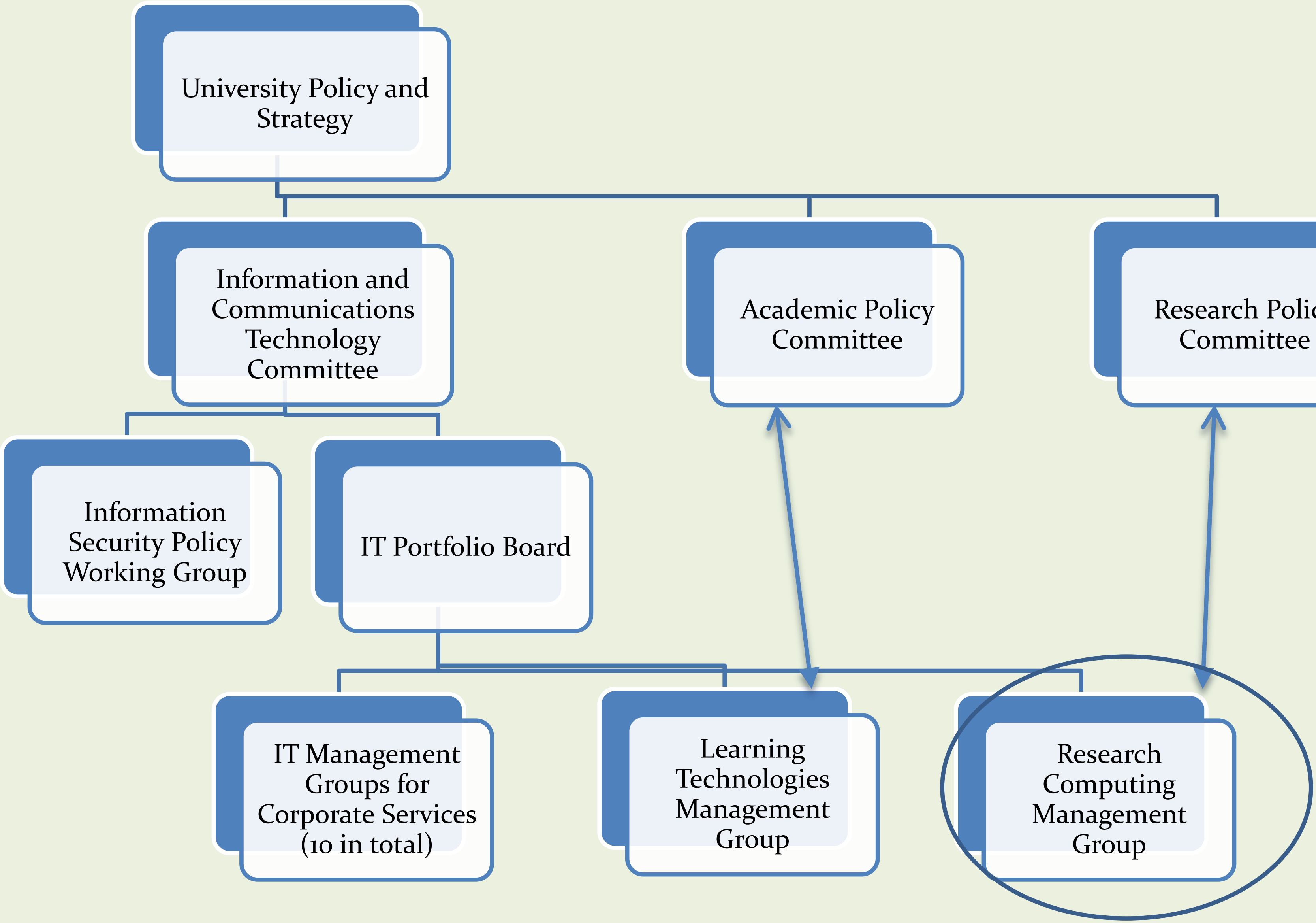
- *Cultural differences*
  - Recognise different cultures and mind sets
    - research community and IT specialists in central services
    - different professional language, expectations and working practises
    - management of a research project usually requires a different, iterative methodology 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

Top Tip: how to get researchers' attention?

- Research grant pre-award costing (LUCRE)
  - 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

Add governance....

Governance



University Policy and Strategy

Information and Communications Technology Committee

Academic Policy Committee

Research Policy Committee

Information Security Policy Working Group

IT Portfolio Board

IT Management Groups for Corporate Services (10 in total)

Learning Technologies Management Group

Research Computing Management Group

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
  - E.g. Sharepoint for 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)