University of Leicester — IT Services

JISC · The University of Nottingham · wellcome trust

HALOGEN

History Archaeology Linguistics Onomastics and GENetics: The Roots of the British interdisciplinary research database

The Halogen project is part funded by JISC in the Managing Data Research programme to implement a cross disciplinary research database to support the Roots of the British collaboration at the University of Leicester. **www.le.ac.uk/halogen**

Partnerships / People

The Roots of the British is a collaboration across a number of departments and researchers at the University of Leicester and other institutes. Key team members are listed below.

Project Manager

- David Carter, IT Services

Team members

- Dr Jonathan Tedds, Senior Research Liaison Manager, IT Services
- Professor Mark Jobling, Senior Research Fellow in Genetics
- Dr Jo Story, Senior Lecturer in Early Medieval History
- Dr Ian Whitbread, Archaeology and Ancient History (ULAS)
- Dr Phillip Shaw, Lecturer in English Language and Old English
- Dr Jayne Carroll, co-director of the Institute for Name Studies, University of Nottingham
- Dr Turi King, Department of Genetics
- Dr Andrew Bradley, Department of Geography
- Dr Dan Pett, British Museum/UCL

Background and Objectives

We aimed to test the integration of data management planning within the research management process and the advisory services that need to be available through the role of 'IT Research Liaison Manager'.

Purpose

- Implement a cross disciplinary database for a collaborative research project.
- Establish organisational best practice in terms of data management planning and infrastructure for these types of initiative.

Planned Outputs

- A **Requirements Analysis** for the research data management infrastructure required.
- A **Data Management Plan**.
- Data Management Infrastructure Design.
- An Implementation Plan for the Research Data Management Infrastructure.
- A **scalable research data management infrastructure** to support HALOGEN from its pilot phase to its end goal of becoming a nationally available source of information.

Research

ROOTS of the BRITISH

Histories, Genetics and the Peopling of Britain. 1000 BC – AD 1000

The fundamental population history of Britain and the roots of the identities of the historical nations of the island (the Welsh, Scots and English) is a contentious subject, especially regarding the legacies of the 'ancient peoples' (known canonically as the **Celts, Romans, Angles, Saxons** and **Vikings**) who are said to have migrated into it, peacefully or as invaders, by the end of the first millennium AD.

The longest-established scholarly traditions – history and philology (the history of language) – have traditionally emphasized mass migration as the fundamental mechanism for major cultural phenomena such as the origins of Celtic languages or of Anglo-Saxon kingdoms in Britain, and to a large extent continue to do so. Modern popular accounts often unquestioningly accept such explanations of the past as established fact. However, many archaeologists reject mass migration as the default explanation for such processes. Genetics potentially offers an entirely independent source of evidence and methodology for addressing these issues, and some present it as a 'magic bullet'.

Two geographical areas are chosen for the HALOGEN pilot: West Lancashire and Wirral/Cheshire, alongside Leicestershire/Warwickshire. Both contain transitions from high to low degrees of Scandinavian cultural influence, as evidenced in place-name densities.

For the Western region, the influence is from Norse Vikings who settled from Ireland at the beginning of the 10th century, while the Central region spans the boundary of the Danelaw, the region under the administrative control of Vikings of Danish origin from the late 9th century.

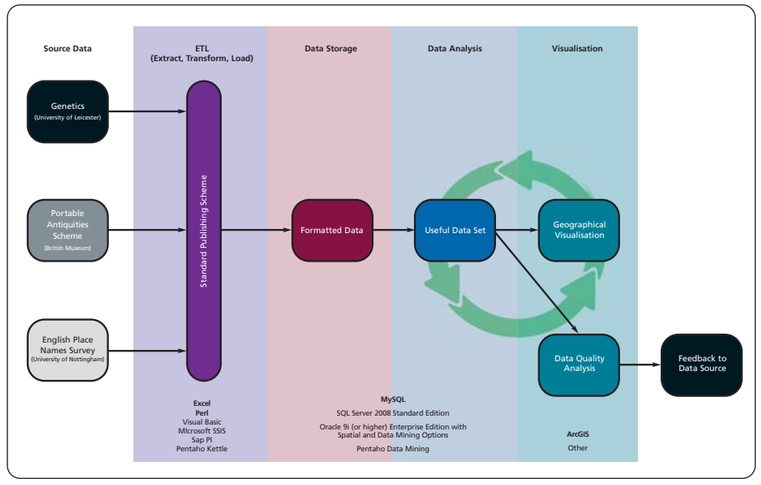
We aim to ask whether the transitions we see in place-names are mirrored in the other datasets we have assembled within HALOGEN.

As well as historical questions, methodological questions are addressed:

- What is the best way to code geographical locations across different data types?
- What geographical units should we use as classifiers: current counties, pre-1974 counties, parishes, poor law unions?
- What is the impact of different geographical resolutions in the different data types?
- If we use interpolation methods to infer data at non-sampled sites, which ones are most suitable to the different data types.

HALOGEN Research Data Management Model

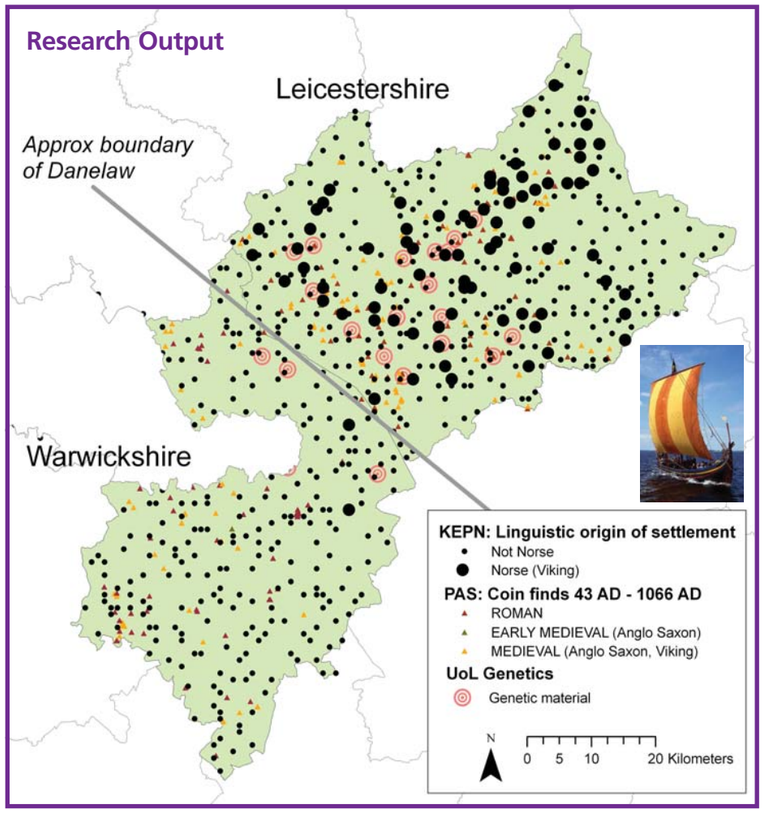
The generic database management model identifies that for database requirements like those of HALOGEN there are five stages or processes that need to be considered.



- **Source Data** – This involves obtaining and storing the source data that will be used as input to the research database. Included within this are procedures to 'refresh' data as new versions become available.
- **Extract, Transform and Load** (ETL) – This involves procedures to extract from the source data those items which are of relevance to the research group. In many cases the raw data may need to be cleaned, formatted in some way in order to improve its integrity or make it compatible with other source data sets and with visualisation and analysis tools. For example, it may be useful to introduce common codes for 'regions' or check that the entries in specific fields are complete and accurate. Finally, there needs to be a way of loading the required data into some type of database.
- **Data Storage** – This involves procedures to store and manage the core data required by researchers. For example, procedures and policies will need to be defined for governing the backup, recovery and access of data.
- **Data Analyses** – This involves applying various tools and techniques to the data to produce information. In some cases it may involve extracting selected information for analyses using tools like SAS, SPSS or R.
- **Visualisation** – This is one specific form of data analyses. From a HALOGEN perspective the geographical visualisation of data is the key user requirement and so it is appropriate to consider it as a separate stage.

The diagram also identifies some of the many software tools that could be used to support different stages of the model. Each institution needs to choose those products which best suit their requirements and the skill sets of those involved in any project.

Research Output



Key Lessons Learned To Date:

Technical/Process

- **Data licences and confidentiality issues** – Need to establish any 'constraints' on using third party data early in the life of a project. This can drive specific requirements. For example, the British Museum/Portable Antiquities Scheme had confidentiality issues relating to the geographic location of finds and this led to the need to 'obscure' data.
- **Cleaning and formatting data** – Think about the analyses tools you will use. Do they require data to be represented in a specific way, what format and content will they support?
- **Establishing a common geographical reference** – How do you generate this from diverse datasets with different levels of spatial granularity. Specialist Geographers need to be involved in this type of project.
- **Selecting tools, database and visualisation technologies** – Investigate cost, skill sets, availability and portability. Leverage experience within the organisation and any 'campus wide' license agreements to reduce the risks and costs of your project.
- **Development methodology** – Prototyping is key. Researchers often find it difficult to articulate true requirements. Deliver a series of 'working examples' to users to drive out real needs.

Organisational Challenges

- **Cultural differences:** Recognise different 'cultures' and 'mind sets' exist between research community and IT specialists in central services. Different professional language, expectations and working practises. Project Manager needs to consider and manage this.
- **Research Liaison Role:** An IT specialist with strong research background enables effective ways of 'liaising' with research community and bridging any gaps in understanding.
- **Leveraging expertise within the organisation:** Look for 'specialists' across the organisation that can help with your problem. For example, in HALOGEN expertise from Geographers and Astronomers was key.

Tedds J.A., Carter D.P. *et al.* email jat26@le.ac.uk