

**Notes and action points from a meeting of BSTM or Thesaurus of British Surnames (TOBS) Steering Committee on 6 February 2002 at PRO**

Present:

Steve Archer (Federation of Family History Societies, National Burials Index);  
Peter Christian Goldsmiths College);  
Louise Craven (Authority Files Editor, PRO);  
John Dawson (University of Cambridge, Computing Service);  
Ian Galbraith (Origins.net, Chairman [chair]);  
Ben Laurie (Free BMD);  
Kevin Schurer (University of Essex, Department of History; Director, UK Data Archive);  
David Thomas (Head of e-Access, PRO)

Apologies: Chris Kutler, (Systems Developer, e-Access, PRO)

- It was agreed that a website for the project would be established. Kevin to make this active by Monday 11<sup>th</sup> February; URL to be [www.data-archive.ac.uk/surnames](http://www.data-archive.ac.uk/surnames)

**KS**

- Louise to amend PID as agreed (see attached). Further sections of the PID not ready for completion until the issues of funding and the initial phases/outcomes have been agreed.

**LC**

- Work to begin on a funding bid in order to clarify the issues and expectations of the group: Kevin to investigate Marc Fitch funding/resources and Ben Laurie will investigate the possibilities of geneology.com.; Ian to contact GSU.

**KS/BL/IG**

- Initial stages of the project agreed to be
  - i. a collection of sample data (and knowledge base accompanying)
  - ii. surname frequency tables
  - ii. definition of the data structure
  - iii. the creation of a typology as a first step to defining a variantand
  - iv. the database defined

**ALL**

*Contributions toward i.:*

A sample dataset is to be developed from the databases/info held by the organizations at 2.4.

In this context:

- Kevin to make available the dataset derived from the 1881 Census (provenance the LDS) used for his work at the Science Museum which has surname, frequency by parish and frequency by county (taken from place of enumeration).

**KS**

- Ian to approach formally those organizations whose representatives were not at today's meeting, including GSU.

**IG**

- It was suggested that Margaret Spufford be approached for dataset of names from the Hearth Tax Project

**KS or DT**

- PRO was asked if 1901 Census data/subset is available. Copyright issues are involved here. Louise/Chris Kutler to investigate

**LC/CK**

- It was agreed that the availability of datasets and the result of funding investigations be publicized via the web site.
- Next meeting: **Thursday 21<sup>st</sup> March at 2pm**. Venue TBN. Funding, progress with sample data sets and with ii and iii to form main items for agenda.

slmc  
13/2/02

**Added from Objectives from IG's original document**

- *process specific datasets to create an initial set of surname groups*
- *create a standardised technique for the lexical decomposition and analysis of British surnames*
- *develop suitable algorithms for automated grouping of surnames*
- *define a standard format for identification of surname groups*
- *define a standard format for recording sources of surname variants*
- *define a means of dealing with surnames which may be members of multiple groups*

**create an online (web-based) database containing the thesaurus]**