

Population And Migration Statistics (PAMS) Committee (Scotland)

Population and Migration Statistics Group

Scottish Information and Linkage Collaboration (SILC)

Introduction

1. This paper provides a high level introduction to the Scottish Informatics and Linkage Collaboration (SILC), the collaborators involved in SILC and the shared services used by SILC. It also provides further detailed information on Information Governance and the role of the National Records of Scotland (NRS) Indexing Team.

Vision for SILC

2. SILC is a collaboration between many academic and public bodies to ensure that Scotland realises the benefits that can be derived from legal, ethical and carefully controlled use of administrative, survey and other types of data. SILC has been created to ensure that there is strategic coherence across initiatives which require the secure processing of individual level data and/or data linkage. SILC is an overarching structure that enables provision of a national research data service for Scotland.
3. SILC and its partners are committed to ensuring that citizens' privacy is maintained whilst simultaneously ensuring that research which is in the public interest is supported, allowing for the benefits of multidisciplinary and multi institutional work to be realised (both within public policy and economic development) through the provision of effective and efficient infrastructural provision.
4. SILC operates in-line with the Guiding Principles for Data Linkage.

Current Position

5. SILC is now fully operational and open for business. A number of projects are at varying stages in the process from Farr, ADRC and Scottish Government. We are looking at methodological improvement around linkage as well as progressing read through arrangements (i.e. where we maintain the linkages between data sets that are held separately but can then be efficiently linked within SILC when the appropriate conditions have been met). We are also working with partners in the Census team, Scottish Longitudinal Survey (SLS) and the Scottish Health and Ethnicity Linkage Study (SHELS) to increase the amount of data that will be available via SILC.
6. So far projects have all come from academia and the central public sector. The SILC team are keen to discuss any ideas for linkage projects from the wider public sector or the third sector that could be developed. We continue to

experience long lead times around data sharing ahead of analysis and are prioritising how we can improve this.

7. We would encourage project proposals (no matter how early in development they are) as we will be able to help to progress these.

SILC Management Structure

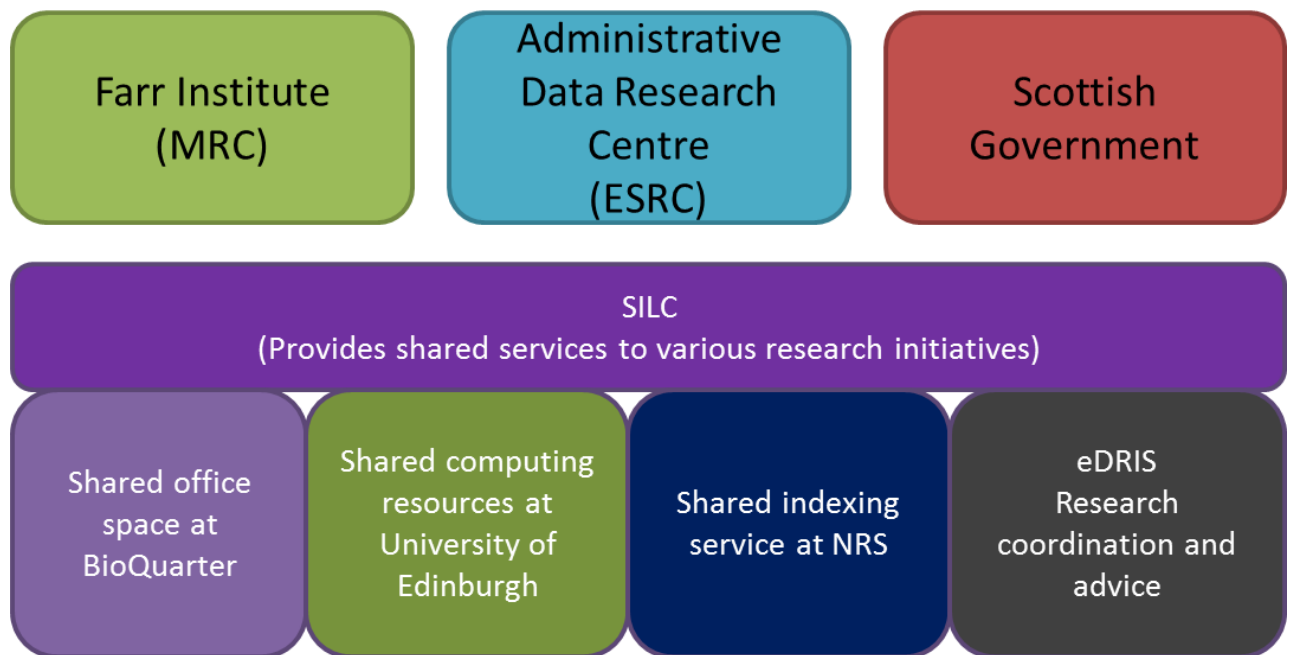
8. The strategic direction for SILC is set by the SILC Strategic Management Board (SMB), which is co-chaired by Tim Ellis and Harry Burns. The SMB reports to the Scottish Government co-ordinated Data Management Board.
9. The SILC Operations Group is responsible for ensuring that SILC operates in line with the priorities of the partner organisations and that the SILC shared services are delivered in an efficient and effective way. The Operations Group reports to the SMB.
10. Some functions of the Operations Group are taken forward by a number of short life working groups (e.g. communication and engagement) and these groups report to the Operations Group and in some cases directly to the SMB.

SILC Collaborators and Shared Services

11. SILC was established in March 2014, bringing together Farr Institute @ Scotland, Administrative Data Research Centre (Scotland) and Scottish Government investment (the Data Sharing and Linkage Service – DSLS). We are currently in discussion with other potential partners who wish to join the collaboration and are open to interest from others.
12. SILC collaborators can make use of the following shared services for data linkage projects:
 - Shared accommodation at No.9 BioQuarter, Edinburgh (including secure analysis facilities for visiting researchers)
 - Shared computing resource at the University of Edinburgh
 - Shared indexing service at National Records of Scotland (NRS)
 - electronic Data Research and Innovation Service (eDRIS) (NHS National Services Scotland (NSS)) research co-ordination and advice

Figure 1 below illustrates the organisations involved in SILC.

Figure 1 – SILC collaborators and shared services



Farr Institute

13. The Farr Institute @ Scotland is a collaboration between six Scottish Universities and NHS National Services Scotland. It is funded by the Medical Research Council (MRC) and data linkage projects for the purposes of medical research come in through this route. The Farr Institute reports to the MRC.

Administrative Data Research Centre (ADRC)

14. The ADRC is a partnership between universities, government departments and agencies, national statistics authorities, the third sector, funders and researchers. It is funded by the Economic and Social Research Council (ESRC) and data linkage projects from academics for the purposes of socio-economic research come in through this route. The ADRC reports to the ESRC. There are four ADRC centres, one in each of the four nations of the UK.

Scottish Government (SG)

15. SG has committed funding to SILC to support 30 SG data linkage projects via the DSLS. This work is being co-ordinated by NRS and the Office of the Chief Statistician. SG report to Scottish Ministers.

University of Edinburgh (UoE)

16. Staff in the Advanced Computing Facility (ACF) at UoE are responsible for providing the secure computing infrastructure for facilitating linkage of the variables of interest to the researcher (payload data) received from data controllers using the linkage keys provided by the NRS Indexing Team. UoE also provide accounts to facilitate secure access to these datasets to researchers. UoE never see any names, dates of births, postcode etc.

National Records of Scotland (NRS)

17. NRS is responsible for providing the indexing service. As part of this, NRS receive personal identifiers (names, dates of birth, postcode etc.) for each of the datasets to be linked. NRS replace this information with randomly generated linkage keys which are returned to each data controller. A different key is sent to each data controller, with a master lookup file sent to UoE to allow the datasets to be linked.

NHS National Services Scotland (NSS)

18. NSS is responsible for providing eDRIS research co-ordinators who provide support to the researcher throughout the process of a data linkage project.

Information Governance

19. The relationships between the SILC shared services will be governed by a series of formal contracts setting out the roles of the partners. SILC isn't a stand-alone entity so the contractual position is complicated but the excellent partnership working allows us to present a simple entry route to researchers.
20. Data linkage projects are required to obtain both ethical and public interest approvals. It is up to data controllers to ensure that all the correct conditions are met but they are supported by eDRIS who work within local arrangements to facilitate safe research. For example where health data is involved the projects will be considered by the Public Benefit and Privacy Panel (PBPP) who approve projects, ADRC projects are approved by the national ADRC approval board.
21. Data exchanges between data controllers and researchers are governed by data sharing agreements.
22. The NRS Indexing Team acts as a data processor on behalf of the data controller. This data processing is governed by data processing agreements.
23. These arrangements are all overseen by the SILC board and put into action by the Operations Group.

Data flows

24. Figure 2 shows the overall indexing and linkage process. For simplicity, the figure does not cover the approvals process which is required before linkage can occur. The validation and quality assurance processes are not shown, to help improve readability, but the NRS Indexing Team will carry out quality assurance on files it receives and creates.
25. The NRS Indexing Team uses the following process to generate the files which allow the datasets to be linked
 - Each data controller agrees their data set can be linked for the research project.

- The data controller provides a file containing only the Personal Identifiable Information (PII) including a unique SourceID from their data set. The PII typically includes names, dates of birth, postcode etc. The SourceID can be the RealSourceID but if this has any intrinsic value in itself, such as Scottish Candidate Number (SCN), it should be replaced by a generated neutral SourceID. They do not send the payload data to the indexing service.
- This data is then matched to the Research Indexing Spine by the Indexing Team who provide them back a source key that contains a IndexedSourceID and corresponding SourceID which they store securely in their own system for the duration of the project.
- This is repeated for each data source with a different IndexedSourceID being created for each data source for reasons of security.
- The indexing team also create a Master Index File which includes a master ProjectID and the IndexedSourceIDs only and pass this to the linking team.
- The indexing team then delete all the datasets they have created.
- The linking team securely store the Master Index File (for the duration of the linkage project).
- New IndexedSourceIDs are created for each project to prevent any linkage of data between projects.

26. The following process allows the keys created by the NRS Indexing Team to be used by the UoE to facilitate linkage of the datasets:

- The data controllers extract the research variables of interest to the researcher (payload data) from their systems (as specified in the data sharing agreement) and replace their RealSourceID and PII with the IndexedSourceID and pass this to the linking team. No other identifiers are required for the linking process.
- The NRS Indexing Team transfer the Master Index File to the linking team (ACF at UoE), who use it to replace the IndexedSourceID on the dataset with the ProjectID.
- This process is repeated for all source data resulting in a set of data sets including only the information to be analysed all with a common ProjectID.
- These data sets can now be linked together by the researcher using the statistical analysis tools available in the safe setting.

Figure 2 – Data Linkage process separation of function

Data Linkage without Read Through

