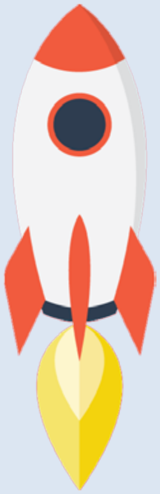
PolicyWise
for Children & Families



SAGE Data Access Manual

Version 1.4

Revision History

Version	Date	Produced by	Change
Jan 26	January 2016	Amanda Lau	Draft items to be included in Manual
V 1.0	June 16 2016	Amanda Lau	Draft with full structure of Manual
V 1.1	June 24 2016	Amanda Lau	Major revisions
V 1.2	June 30 2016	Lucie Richard	Formatting and minor revisions
V 1.3	October 2 2017	Hannah Lloyd-Jones	Minor revisions and changes to review process
V 1.4	November 23 2017	Hannah Lloyd-Jones	Minor revisions following SAGE Implementation Meeting discussion

Contents

Revision History 2
Section I: Introduction 4
Overview 4
Why access data through SAGE? 4
Section II: Data Access Process 5
Figure 1- Procedure overview 5
Step 1: Secondary data user submits Data Inquiry Form 6
1.1. Eligibility 6
Step 2: SAGE staff consults with Secondary data user 6
Step 3: Secondary user data submits Data Access Request form 6
Step 4: Data Access Request Review 7
4.1 Review Criteria 7
4.2 Review timeline 7
4.3 Review outcome 7
Step 5: SAGE staff provides a preliminary cost estimate 8
Step 6: Secondary data user applies for ethical approval from a recognized research ethics board 8
Step 7: SAGE staff and Secondary Data User sign Data Access Agreement 8
7.1 Intellectual property 8
7.2 Repatriation of derived variables 8
7.3 Privacy of Participants 8
Step 9: Secondary data user completes analysis by accessing customized de-identified dataset, analytic tools, and research supports through SAGE'S Analytic Environment. 9
8.1 Alternate methods of data access 9
8.2 Research Output 9
Step 10: SAGE evaluates research impact and tracks publications 10
Contact Information 10

Section I: Introduction

Overview

SAGE (Secondary Analysis to Generate Evidence) is a platform where research data, service delivery data, and (at a later stage) administrative data are catalogued and managed for secondary research and policy use. SAGE provides technical infrastructure and governance processes that protect participant/client privacy and ensure ethical re-use of data. Our goal is to improve outcomes for Alberta's children, youth, families and communities by enabling new uses of data to inform social, educational and health policy, practice and service planning.

Why access data through SAGE?

Millions of dollars are being invested in data collection in both the academic and public domain. According to a report from the Canadian Research Data Summit, many valuable research data are being "under-utilized and […] important publicly funded assest[s] are being wasted". While true for research data, this under-utilization can also be seen for data collected by service delivery organizations and within government sectors. Within these sectors, data is collected primarily for the provision of services – however, there is a wealth of knowledge to be discovered through its re-use.

Re-use of existing data beyond its original purpose has numerous benefits. These include:

- Replication and verification of results, enabling greater transparency
- Increased collaborative opportunities with the original data collector
- In instances were data linkage is possible, datasets from multiple sources can be combined to answer more complex question and potentially engage more interdisciplinary expertise
- Reduced burden on client/participant by avoiding data collection duplication
- Increased return on investment

SAGE provides a number of value added services that optimizes the use of existing data. For one, SAGE evaluates and enhances data quality in partnership with the original data collector to increase re-use potential, which can be beneficial for secondary data users. SAGE also ensures that datasets have comprehensive documentation and metadata (data about data), so that the studies and their variables are easy to understand, thus allowing for easier and appropriate re-use. Second, SAGE is leveraging PolicyWise' expertise in secondary research, data linkage, and analytics, lending advice to secondary data users were possible. Most importantly, SAGE does not only build bridges to new research, but facilitates the building of partnerships – be it amongst researchers, policy makers, or community practitioners.

Section II: Data Access Process

SAGE works with Secondary Data Users to ensure easy and feasible data re-use. This document describes the processes and procedures required for accessing data from SAGE.

Access to SAGE data is dependent on, but separate from, funding and ethics. Prior to data release, applicants are required to have ethical approval and encouraged to have funding in place.

Figure 1- Procedure overview

Step 1: Secondary data user submits Data Inquiry Form

Secondary Data Users complete the online Data Inquiry form stating the purpose and potential impact of their research.

1.1. Eligibility

Secondary Data Users should:

- Demonstrate that they or their research team (if applicable) have the content and analytic skills, knowledge, and resources to carry out the proposed project.
- Be able to obtain ethics approval from a recognized ethical review board.
- Propose a project that is for non-commercial purposes.
- Must abide by the "Terms of Use" for the dataset(s) of interest as found on SAGE's Dataverse

Metadata about SAGE data assets is available on SAGE's Dataverse to help guide proposal/hypothesis/aim development. Within the each dataset's metadata, Secondary Data Users are encouraged to pay particular attention to the "Embargo" and "Terms of Use" which outlines availability of variables and access procedures, respectively.

Step 2: SAGE staff consults with Secondary data user

After review of the inquiry, SAGE staff will contact the Secondary Data User to discuss the feasibility of the project. Where requested, SAGE staff can support datasets exploration, identification of relevant variables, ethic submissions, as well as provide methodological and statistical advice.

An important goal of SAGE is to facilitate collaboration and to support research and data capacity building. During the consultation stage, SAGE will discuss potential collaborative opportunities between the Secondary Data User and the original data producer (i.e. researcher or service delivery organization). In some instances, collaboration with the original research team is required at various stages of the project.

SAGE will also discuss any dataset specific terms of use, as stipulated by the data producer (for example, form of data producer acknowledgement). SAGE staff will also ensure that the Secondary Data User understands that access to the data will be via the secure Analytic Environment.

Step 3: Secondary user data submits Data Access Request form

After consultation, SAGE staff will send the Secondary Data Users a Data Access Request form. The Secondary Data User will complete and submit the Data Access Request online. The Data Access Request is akin to completing a grant funding application. This request clearly states research aims and hypotheses, as well as describes the relevant exposure, outcome, and confounders to justify the data being requested. A list of study variables will also be submitted as part of the Data Access Request. A budget, as well as applicant and co-applicants academic CVs will be submitted. We recommend that the Secondary Data User work with their co-applicants (if appropriate) to review the Data Access Request before submitting it. Note that if the scope of your research changes, you need more variables or if your research team changes from that stated on the Data Access Request form, we may need to review your request again and sign a new Data Access Agreement.

Step 4: Data Access Request Review

All Data Access Requests are reviewed by SAGE staff and the original Data Producer (or representative) if they have indicated that they wish to participate in Data Access Request reviews. In some cases, one or more external peer reviewer will also be asked to review the request.

SAGE staff review the request for feasibility and appropriate use of data. The Data Producer and the external reviewer(s) (if appropriate) ensure that the proposed research is relevant, sound, and consistent with the data being requested. Note that postgraduate research students need to submit a copy of their committee's thesis approval as part of their Data Access Request.

4.1 External Review Criteria

If an external review is required, each reviewer evaluates and provides comments for the research proposal based on the following criteria:

- Overall Impression
- Background, objectives, potential contribution to knowledge
- Approach and Methods
- Impact and Knowledge Translation
- Project Management, Research Team, and Partnerships

Additional criteria (does not count towards overall score)

- Budget

The detailed reviewed criteria can be found in the *SAGE Access Review Criteria* document.

4.2 Review timeline

Review processes take approximately 4-6 weeks.

4.3 Review outcome

After evaluation, SAGE staff will inform the Secondary Data User of the review outcome, as well as provide a copy of comments/suggestions offered by the reviewers. Unsuccessful applicants may revise and resubmit proposals. These Secondary Data Users are advised to speak to SAGE staff before resubmission.

Successful applicants and the title of their proposed research may posted on PolicyWise' website.

Step 5: SAGE staff provides a preliminary cost estimate

Successful applicants may be charged an access fee based on a cost-recovery model (see *SAGE cost recovery guidelines*). These fees are determined on a project-to-project basis, and are required to reimburse operational costs such as processing of data requests and use, as well as, data preparation and extraction. This fee may be waived in certain situations such as undergraduate research project or for service delivery organizations.

Step 6: Secondary data user applies for ethical approval from a recognized research ethics board

Prior to data release, secondary research proposals need to have ethical approval from a recognized ethics board (i.e. institutional or community). Approval can be sought during or after the review process.

Step 7: SAGE staff and Secondary Data User sign Data Access Agreement

SAGE and Secondary Data Users sign a Data Access Agreement that outlines the terms and conditions for provision of services and data.

Some of the terms and conditions outlined in Data Access Agreement include:

7.1 Intellectual property

SAGE does not claim any intellectual property rights with regards to the results, discoveries, invention, or work that could stem from research where the data was used.

7.2 Repatriation of derived variables

Derived variables are variables constructed by the Secondary Data User during the course of a proposed secondary research project. These variables may include new data (where applicable) or derived data created from existing data.

To enrich the SAGE research platform, Secondary Data Users are required to return the results of the secondary analysis (i.e. derived data or new data), along with their appropriate documentation, back to SAGE. The exact nature of the results to be returned will be determined in collaboration between the Secondary Data User and SAGE. The derived variables will be incorporated into SAGE data assets for use by other Secondary Data Users. This will allow researchers to increase the value of data assets and build upon previous research.

7.3 Privacy of Participants

Approved Secondary Data Users agree to use and manage data from SAGE with strict confidentiality. Secondary Data Users agree that data will not be used to establish the individual identities of any of the participants.

The following documents should also be included with the Data Access Agreement: secondary data research proposal, list of requested variables, list of research team, and a copy of the ethic approval certificate. Once the data access agreement is complete a request for new users is submitted to the SAGE IT manager. Also appropriate software is requested and deployed on the SAGE environment.

Step 8: All members of research team attend a SAGE orientation session and sign Code of Conduct Acknowledgement.

All new users attend a SAGE orientation session, before signing a Code of Conduct Acknowledgement and obtaining access to the SAGE Analytic Environment.

Step 9: Secondary Data User completes analysis in the SAGE Analytic Environment

The Secondary Data User will access their customized, de-identified dataset through the SAGE Analytic Environment. This Analytic Environment is a secure remote computer system that acts as a personalised access point to data and analytic tools (SPSS, STATA, SAS, Tableau, etc.) so that Secondary Data Users can store and work on their datasets. Since the environment is moated, there is no internet access and data cannot be removed from the environment. This type of environment limits the risk of privacy breaches, while facilitating multidisciplinary, multi-organizational collaboration.

In rare instances, Secondary Data Users may be required to access data via a physical enclave or secure file transfer protocol (SFTP). These methods will depend on the sensitivity of the data, and are up to the discretion of SAGE.

8.1 Alternate methods of data access

- Physical enclave

  Very sensitive data will requires access through a physical enclave. Enclave spaces are located in PolicyWise offices in Edmonton and Calgary.

- Secure file transfer protocol

  A data file can be securely and electronically transferred to the Secondary Data User. Secondary data users will be assigned a username and password when they first access the dataset.

8.2 Research Output

As part of the process to safeguard the privacy of research participants during secondary use of data assets, SAGE staff review (or “vet”) all research outputs to be shared externally, including the following permitted outputs:

- Syntax documents (programs and/or logs of programs);
- Descriptive or Results tables;
- Figures

Please note that SAGE does not vet the accuracy of programs or outputs.

Step 10: Evaluation of research impact and tracking of publications

SAGE staff should be informed when research is complete, and access to data is no longer required. The period of data access is study specific and described in the Data Access Agreement. However, this period may be extended by the Secondary Data User through written request.

All research output (i.e. manuscripts, presentations, abstracts, reports) should be submitted to SAGE prior to dissemination for information purposes only. Information obtained from these documents will not be used or disclosed without the written consent of the authors.

In efforts to highlight the benefits of data sharing, SAGE encourages publication and impact tracking. Upon project completion, Secondary Data Users are asked to provide a Final Research Report bi-annually, for up to three years. This report aims to summarize the research output and knowledge mobilization resulting from the secondary research project. A template of the Final Report can be found on the PolicyWise website.

Contact Information

If you have any questions regarding SAGE's Access Manual, please contact SAGE staff at data@policywise.com.