

Secondary Analysis to Generate Evidence (SAGE)

Developmental Evaluation Report

Project Team: Kendra Leavitt, Laurie Vermeylen, Naomi Parker, Dr. Cathie Scott

Date: 10 September 2017



Policy Wise
for Children & Families

Primary Contributors

Kendra Leavitt, Naomi Parker, Laurie Vermeylen, Dr. Cathie Scott

Acknowledgements

It is with thanks we acknowledge the many individuals and organizations who have contributed their wisdom, experience and perspectives to this project. PolicyWise would like to thank each who participated in the interviews with us.

Suggested Citation

Leavitt, K., Parker, N., Vermeylen, L., Scott, C. (2017). *Secondary Analysis to Generate Evidence (SAGE): Developmental Evaluation Report*. Edmonton: PolicyWise for Children & Families.

Sharing Guidelines

It is the hope of all those who contributed to this project that these findings are shared and used to benefit others and inform policy and practice to improve child, family and community well-being. PolicyWise asks the intent and quality of the work is retained; therefore, PolicyWise for Children & Families must be acknowledged in the following ways:

- In all published articles, power point presentations, websites, signage or other presentations of projects as: *Name of Project led by PolicyWise for Children & Families*
- The PolicyWise logo must be used in conjunction with this acknowledgement in all of the above instances.
- This product and content included in it may not be used for commercial purposes.
- No derivative works and publications. You may not alter, transform or build upon this material without permission.

Key Messages

- SAGE has undergone several significant transformations since its initiation, including a name change, introduction of two Advisory Committees, evolving goals, and a move towards secure remote access.
- Protecting privacy, preventing participant identification, and ensuring consent or ethics forms account for potential data sharing and secondary analysis are the main concerns associated with administrative, research, and community data.
- Established protocols and vetting processes for accessing data mitigates risk and protects sensitive data.
- There is growing recognition of data sharing benefits (i.e., preservation of data, advanced scientific knowledge, improved efficiency, etc.). Education and promotion are facilitators to this cultural shift.
- A number of cultural barriers to data sharing, secondary analysis, and data repositories continue to exist. Many of the barriers are related to data ownership concerns, cost, funding, and frameworks for sharing data nationally and internationally. Creative and practical strategies are being used to address cultural barriers.
- Supporting researchers early and developing clear data sharing processes are goals of SAGE and other like-minded repositories.
- Relationships are a critical element to data sharing. Three types of relationships were highlighted: other data repositories, data producers and accessors, and policymakers or institutions.
- Building capacity among researchers, students, and new academics facilitates a skilled workforce, secondary analysis, and the ability to address gaps related to managing and sharing data.
- Data sharing has the potential to significantly increase impactful work.
- Promotion and visibility are important elements for SAGE's success, with a number of approaches being implemented or explored.
- A number of requirements must be met for data to be deposited or accessed in order to ensure high quality data and protection of privacy.
- Though operations vary by data repository, SAGE's operations generally align.
- Implementing two Advisory Committees is leading practice. Interview participants indicated concerns related to the frequency of meetings and wanting to contribute more substantially, though.

Executive Summary

SAGE (Secondary Analysis to Generate Evidence) is a data repository that operates within PolicyWise for Children & Families (PolicyWise). As part of a commitment to ongoing learning, adaptation and improvement, PolicyWise has conducted an internal evaluation of SAGE. Findings from interviews with the SAGE team, those internally familiar with the initiative, and experts in the field, as well as a document review, were synthesized. The purpose of the developmental evaluation is to:

- Understand and describe the potential outcomes and impact of SAGE;
- Identify features of SAGE or context that may have significant influence on outcomes; and
- Develop and implement a framework for ongoing monitoring, adaptation and improvement, linked with PolicyWise' Impact Assessment Framework.

This evaluation will use a Developmental Evaluation (DE) approach that will facilitate a focus on continuous quality improvement, providing project sponsors with real-time feedback in order to inform ongoing decision-making and adaptation of SAGE. Findings from this evaluation will be used to inform SAGE's understanding of: design and delivery; outputs associated with the implementation of secondary data repository; and projected achievements.

The Canadian Research Data Summit 2011 Final Report identified underutilization of research and data collected in Canada (Research Data Canada, 2011), presenting an opportunity to respond, and contributing to partnerships between SAGE (formerly the CDCA) and others across Alberta. The purpose is to store, clean, catalogue, and manage data for secondary research, policy use, and addressing gaps in data sharing (ACCFR, 2013, pg. 3).

Findings indicate that the field of data sharing is rapidly evolving and expanding. Aspects such as privacy and ethics, legislation, culture and associated barriers, goals, outcomes, and impacts, as well as operations and governance are areas of consideration and advancement. The findings further indicated that SAGE is at the forefront of leading practice and innovation, particularly as it relates to eventual plans of linking administrative, research, and community data.

Privacy and Ethics

Privacy and ethics were a substantial point of discussion in all research components. It is clear that as data sharing grows, more attention and consideration is being given to protecting privacy and ethics processes. While administrative, research, and community data are similar in many ways, consideration for each of the different types of data should be given. Legislation and regulations around data sharing are different cross-jurisdictionally and between different data repositories. According to interview participants, Alberta legislation related to data sharing is open to interpretation. The majority of those interviewed felt that work needed to be done to facilitate the development of clearer legislation. The goal of data risk management is to maintain privacy while allowing access to meaningful data. Consistently, it was expressed that risk could not be entirely eliminated, but could be effectively managed.

Broader Culture of Data Sharing and Repositories

Cultural changes to data sharing are occurring, including recognized benefits and increased education; however, some barriers still exist. Cultural barriers to data sharing include: reluctance of researchers to share data, lack of knowledge and understanding of data repositories and secondary data use, costs of producing, accessing or depositing, and housing data, and funding constraints.

Increasingly, in this section the research pillar of data sharing was emphasized in the data collection. Though the community, and to a certain extent, administrative pillars are represented, many of the perspectives are shared by researchers in those respective fields, further contributing to an emphasis on the research pillar. This reflects an area for further consideration.

Goals, Outcomes, and Impacts about Data Sharing

Three types of relationships were discussed by interview participants: other data repositories, data producers and accessors, and policymakers or institutions. Building relationships was emphasized across interviews, the document review, and to a lesser extent the literature. As Humphreys (2006) notes, without partnerships long-term preservation and access to data would not be possible; however, in order to build partnerships we must understand the research process and areas or gaps that can be addressed through data preservation and access.

Strengthening and mobilizing capacity was not a theme discussed in the literature; however, it was a goal of most interview participants. Due to the wide range of interview participants building capacity occurred through several different avenues such as research production, data sharing, answering new research questions using existing data, or by preserving and linking data.

According to the OECD (2007), international frameworks for data sharing require further improvement; however, global scientific databases are growing, increasing data sharing and access which:

- Reinforces open scientific inquiry;
- Encourages diversity of analysis and opinion;
- Promotes new research;
- Makes possible the testing of new or alternative hypotheses and methods of analysis;
- Supports studies on data collection methods and measurement;
- Facilitates the education of new researchers;
- Enables the exploration of topics not envisioned by the initial investigators; and
- Permits the creation of new data sets when data from multiple sources are combined (p.10)

Operations and Governance

Operational themes include promotion and visibility, requirements of the data, requirements for accessing data, providing support to data producers and accessors, remaining up to date on leading practices in data sharing and repositories, and sustainability.

The majority of other data repositories had a committee or council, which varied in size and representatives; however, interview participants stated that committee members should be carefully considered, and decisions should be made regarding whether committee representation should be research or process heavy. Interview participants also recommended including ethicists and legal representatives, something SAGE did early on.

Conclusion

Overall, interviews, the document and literature review indicate that SAGE has been working well. Several interview participants stated that SAGE was a relevant player in the field of data repositories, and the eventual ability to link research, community, and administrative data, along with a focus on children and families would make SAGE unique. Interview participants noted that the data already held with SAGE is clean and relevant. Additionally, SAGE was viewed as experts in the area of data privacy.

Table of Contents

Key Messages.....	3
Executive Summary.....	4
Introduction	8
Project Approach: Developmental Evaluation.....	8
Methods.....	9
Findings	10
Background and Evolution of SAGE	10
Privacy and Ethics	11
Broader Culture of Data Sharing and Repositories.....	15
Goals, Outcomes, and Impacts about Data Sharing	20
Operations	25
Governance.....	31
Opportunities and Future Plans.....	33
Conclusion.....	35
References	37
Appendix A.....	39
Appendix B.....	42
Appendix C.....	44
Appendix D.....	47

Introduction

SAGE (Secondary Analysis to Generate Evidence) is a data repository that operates within PolicyWise for Children & Families (PolicyWise). As part of a commitment to ongoing learning, adaptation and improvement, PolicyWise has conducted an internal evaluation of SAGE with the goal to:

- Understand and describe the potential outcomes and impact of SAGE;
- Identify features of SAGE or context that may have significant influence on outcomes; and
- Develop and implement a framework for ongoing monitoring, adaptation and improvement, linked with PolicyWise' Impact Assessment Framework.

This evaluation will use a Developmental Evaluation (DE) approach that will facilitate a focus on continuous quality improvement, providing project sponsors with real-time feedback in order to inform ongoing decision-making and adaptation of SAGE. Findings from this evaluation will be used to inform SAGE's understanding of: design and delivery; outputs associated with the implementation of secondary data repository; and projected achievements.

This report describes SAGE and the broader culture in which data sharing and repositories exist. It further outlines common goals, outcomes and impacts, operations and governance to identify opportunities and future considerations for SAGE.

Project Approach: Developmental Evaluation

For this internal review, PolicyWise has used a DE approach. DE is a process best used for initiatives with high complexity and where innovation is desired. The DE strategy, unlike the more linear process associated with traditional evaluation (i.e., data collection → analysis → recommendations → implementation), allows decision-makers to engage in an iterative process that allows for ongoing learning and course correction (Quinn Patton, 2006). The DE approach is meant to provide an examination of current practices while identifying potential changes that can be implemented. It is meant to promote the strengths of the program while identifying opportunities for improvement. With SAGE in its early implementation, this is a key time to implement new ideas and recommendations.

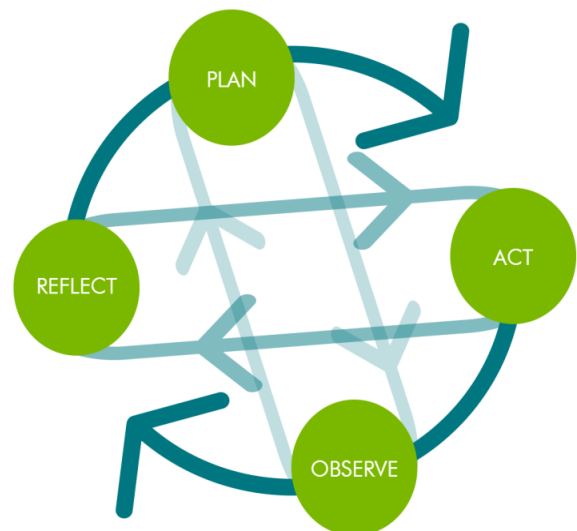


Figure 1: Developmental Evaluation Process

There are four stages to DE that act together to improve services or projects by implementing changes, see Figure 1 (Kemmis, McTaggart, & Retallic, 2004). These four stages are fluid, can happen in various orders, and are overlapping. Each stage is described below:

1. Plan: This involves designing steps towards action to improve on what is already happening. Planning is strategic because it takes into account the constraints of the current context while capitalizing on other expertise and evidence.
2. Act: To implement the changes that are meant to improve upon the program or project as an 'idea-in-action' to determine how changes will roll out within the current context. Action is dynamic and requires fluidity.
3. Observe: To observe the effects of the action in the context in which it occurs and to document the changes, barriers, and constraints.
4. Reflect: To examine the effects as a basis for further planning or action, trying to make sense of the context, the change, and the outcomes.

Methods

Six interviews were conducted with SAGE staff and those internally familiar with the initiative. (Appendix A) Participants were identified by the SAGE team. These interviews were done either over the phone or in person, were audio-recorded, and lasted approximately one hour. Notes were made from the recordings and they were coded separately by two team members and themes were agreed upon.

Throughout the internal interviews, key thought leaders and experts in the field were identified. A list of potential interviewees was developed, capturing perspectives from research and community data and the operational advisory committee as well as representatives from other data repositories (Appendix B and C). Eleven people were invited for an interview via email and nine interviews were conducted. These were done over the phone, lasted approximately one hour, and were audio-recorded. Again, notes were made from the recordings and they were coded separately by two team members and themes were agreed upon.

Relevant documents were identified by the SAGE team and reviewed for key themes. Recent literature was also reviewed, with relevant articles selected based on a previous literature review and search approach completed by the SAGE team. Since the literature review was recently conducted by SAGE a full literature review was not completed.

Findings were then combined from all data collection components to identify cross cutting themes.

Findings

Six main themes based on the interviews, and document and literature review are described below.

- Background of SAGE and evolutions that have occurred since its initiation.
- Privacy and ethics including consideration for research, community, and administrative data, data sharing legislation and agreements, and managing risk.
- Broader culture of data sharing and repositories is then discussed based on culture changes currently happening, cultural barriers that still exist, and the specific ways change is being supported and addressed.
- Goals, outcomes and impacts for data sharing are discussed. This includes how data repositories and researchers are or plan to prioritize data sharing processes, build relationships, strengthen and mobilize capacity, and increase productivity and impactful work.
- Operations and governance of other data repositories and SAGE are then considered.
- Areas that are working well and areas for improvement for SAGE and others, as well as potential next steps and considerations for SAGE moving forward.

Background and Evolution of SAGE

A significant amount of growth and learning has happened since the planning and initiation of SAGE. This section explores the contextual factors that contributed to SAGE's launch and specific evolutions that occurred throughout the learning process.

The Canadian Research Data Summit 2011 Final Report identified an underutilization of research and data collected in Canada (Research Data Canada, 2011) despite research trends indicating that data related to 'child and youth development, health, and well-being' were being collected by governments or generated by researchers in large quantities (ACCFCR, 2013, pg. 4). These reports led to an identified need for increased access to and sharing of existing data in Canada. In response, the Child Data Centre of Alberta (CDCA) was developed in partnership with the former Alberta Centre for Child, Family and Community Research (ACCFCR, now PolicyWise), the University of Alberta Women and Children's Health Research Institute (WCHRI), the University of Calgary Alberta Children's Hospital Research Institute (ACHRI), and Alberta Innovates-Health Solutions (AIHS), with the purpose of storing, cleaning, cataloguing, and managing data for secondary research and policy use and addressing gaps in data sharing (ACCFCR, 2013, pg. 3). Interviews with the SAGE (formerly CDCA) team indicated that CDCA was designed to complement other existing repositories rather than duplicate services. Some aspects of CDCA were modelled after CDCA-like initiatives (i.e., ICES, Population Data BC) that were identified through interviews and scans of other data repositories.

CDCA has adapted and evolved since its initiation, and is described as follows. First, in 2016 CDCA underwent a name change to SAGE, which occurred at the same time ACCFCR also changed its name to PolicyWise. The data repository will be referenced as SAGE throughout the document unless otherwise specified.

Second, SAGE initially had a working group that was later split into two Advisory Committees; the Operational Advisory Committee and the Strategic Advisory Committee. Each has different roles and is arms-length from PolicyWise. This will be discussed further in the governance section below.

Third, SAGE was primarily focused on research data because it generated the most interest and was more streamlined to setup. Since then SAGE has also explored administrative and community data. A goal of SAGE is to store and prepare research, community, and administrative data that can be linked to answer complex or policy relevant questions (ACCFCR, 2013). SAGE team members stated that linking data across the different sectors, particularly administrative and community data, allows for a more holistic understanding.

Fourth, SAGE is moving towards secure, remote access. At the outset, an Enclave room was built to store and access the data. Through the learning process it was determined that such a room would likely not be used, and instead data has been shared through remote access and sending anonymous data. SAGE is continuing to evolve in the dynamic field of data sharing and in response to ongoing learning and technology changes.

Privacy and Ethics

Privacy and ethics were a substantial point of discussion in all research components. It is clear that as data sharing grows, more attention and consideration is being given to protecting privacy and ethics processes. This section will discuss considerations for the different types of research data and addressing privacy or ethics concerns, legislation and data sharing agreements, and managing risk.

Consideration for Different Types of Research

While administrative, research, and community data are similar in many ways, consideration for each of the different types of data should be given. According to interview participants, the greatest concern for administrative data is protecting privacy and preventing identification of individuals due to the highly sensitive information administrative data contains. Discussions with interview participants indicated that some individuals already have the expectation that their data are integrated and shared across different government departments and services. They are less concerned about data sharing and more concerned about ensuring they are not individually identifiable and their information is being used for research purposes only. There are a number of ways to protect individuals from being identified as described by interview participants, including restrictions around who can hold identifiable administrative data, referred to as custodians or prescribed entities. Some of these entities restrict who has access within their organization to identifiable data or only holds de-identified data that is linkable through numerical codes. Interview participants also identified legislation, particularly in Ontario, that requires data to remain with the prescribed entity, meaning that it cannot be sent out and must be accessed onsite or through secure VPN's. Nonetheless, provincial regulations related to administrative data storage and secondary use can be open for interpretation and restrictive according to interview participants. SAGE is

addressing administrative data concerns by working with the Office of the Information and Privacy Commissioner of Alberta (OIPC) and learning from the Child and Youth Data Lab (CYDL)¹, which has access to data from different ministries in Alberta. They are also learning from other models and adapting those processes to the Alberta context.

One interview participant raised an interesting question related to administrative data ethics and the potential for linking data that results in actionable findings. The participant gave an example of linking data that reveals medical treatments which could benefit an individual; however, many consent forms state information will not be returned to participants. This raises questions about harm that could be caused by withholding or providing information to individual research participants and if policies governing next steps should be implemented in response.

Considerations for research data are mostly focused on ethics and consent. According to interview participants, data sharing or linking is often not considered at the beginning of the research project and is therefore not included in the consent forms signed by research participants. As a result, researchers are required to go back to participants for re-consent if they want to share or link their data. Many interview participants described this process as frustrating, especially when the data set is large and participants cannot feasibly be re-consented. Researchers affiliated with universities must also include data sharing or linking in their ethics application. Some interview participants stated that university research ethics boards are often conservative in their interpretation of consent forms and ethics applications, which can prevent researchers from sharing or linking data. These challenges are further complicated when research is cross-jurisdictional. The literature emphasizes a need to develop an ethics review process and frameworks that are cross-jurisdictional to facilitate data sharing and secondary analysis nationally and internationally (OECD, 2007; Dove et al., 2016). The document review showed that SAGE hopes to address some of these issues by developing broad consent language researchers can use to facilitate data sharing, storage, and secondary use (CDCA Operational Advisory Committee Meeting 2, 2015; Proportional Governance and Access Controls Guidelines, n.d.). Additionally, Meslin, Rager, Schwartz, Quaid, & Gaffney et al. (2015), state that conflicts of interest can arise in industry-academic partnerships over data use and ownership. Though financial agreements are important in partnerships they tend to become the focus. The authors suggest increasing attention to other agreement areas such as ethics to avoid conflicts of interest.

“...clients have been generous to give us this as part of the journey, knowing it is going to help service...It’s different coming to a private agency and I feel a huge responsibility to the people we serve to be sure we are being respectful and ethical...I think we have to be sure there is a further step in there to be sure we are, and our processes are not taking advantage of them.”

-Interview Participant

¹ The CYDL is a joint initiative between PolicyWise for Children & Families and participating ministries in the Government of Alberta. The mandate of the CYDL is to link and analyse administrative data from Government ministries, to provide evidence for policy and program development.

The number of community agencies collecting data is growing as agencies look for ways to evaluate their programs and services in an increasingly competitive funding environment. One interview participant noted that community data is important because it has the potential to provide information about services people are accessing, which may not be captured in other research models, such as randomized controlled trials. However, interview participants noted that because community data isn't necessarily used for research, it can be collected in various ways, which can result in "messy and unreliable" data. For this reason publishing can be difficult, making researchers hesitant to use community data for secondary analysis. As interview participants stated, agencies collecting community data are also concerned about privacy, especially if processes to protect privacy are not in place or staff are not trained to collect data properly. Additionally, ownership over community data is still unclear, adding another layer of complexity. For researchers not affiliated with an institution such as university and depending on the type of research, there are a few options within Alberta, including ARECCI, which is not an official ethics board but can assist researchers with protecting participant privacy, and HREBA-CHC, which is an official ethics board that is indicated in the Health Information Act (HIA). Like administrative and research data, SAGE has policies and procedures to protect community research participants, manage ownership of data, and provide access for secondary analysis.

	<i>Key Considerations</i>	<i>Related Findings</i>
<i>Administrative Data</i>	Work with ministries to develop policies and processes to protect data	Privacy and individual identification is the greatest concern for administrative data
	Demonstrate research participant willingness to share data	
<i>Research Data</i>	De-identify data using numerical codes or algorithms even when not required by the data partner	Ethical questions have been raised about actionable research findings and the appropriate response
	Facilitate conversations with other data repositories and governing bodies about potential ethical issues and discuss next steps	
<i>Community Data</i>	Work with researchers early in the research process to ensure proper ethics and consent are obtained for data sharing	Ethics and consent are concerns for researchers sharing data. Consent for data sharing is often not obtained at the beginning of the research process
	Assist with developing ethics and consent submissions that leave open the possibility for data sharing	
<i>Community Data</i>	De-identify community data using numerical codes or algorithms	Community agencies are concerned about protecting client privacy
	Offer support to community agencies by providing training on data collection, analysis, and data sharing	

Data Sharing Legislation and Agreements

As noted earlier, legislation and regulations around data sharing are different cross-jurisdictionally and between different data repositories. According to interview participants, Alberta legislation related to data sharing is open to interpretation. The majority of those interviewed felt that work needed to be done to facilitate the development of clearer legislation. Some interviewees expressed that other provinces such as Manitoba and Ontario have more well-defined data sharing legislation. SAGE is involved in facilitating discussions with different governing bodies as a means to promote data sharing legislation that is in line with the needs of data producers, accessors, and repositories.

Data repositories have their own data sharing agreements between data accessors and depositors. Each repository interviewed had different ways of approaching data sharing agreements, ranging from specific agreements for each project to having one broad agreement for everyone. A number of interview participants stated that having individual data sharing agreements was time consuming and significantly slowed the process down. Having streamlined agreements increased efficiency and prevented the need for further data sharing approvals. Lee, Sung, Barnett, and Norris (2016) argue that strong data sharing agreements facilitate secondary data analysis and good data stewardship. Additionally, data use restrictions are often an unintentional result of poorly drafted agreements from data producers (Kanous & Brock, 2015). Well-articulated and clearly defined data use or sharing agreements not only mediate against ethical issues and ensure compliance of data use, but also prevent inadvertent restrictions and prevent misperceptions related to intellectual property.

<i>Key Considerations</i>	<i>Related Findings</i>
Develop a streamlined data sharing agreement that can be applied to every project while leaving room for adaptation	Data sharing agreements for each project is time consuming and slows the process down
Assist researchers with drafting clearly defined data sharing agreements	Unintentional data restrictions may result from unclear data agreements produced by researchers

Managing Risk

According to interview participants, the goal of data risk management is to maintain privacy while allowing access to meaningful data. The majority felt that risk could not be entirely eliminated, but could be effectively managed. When asked how risk is managed, participants emphasized the need to build trust with data accessors, producers, and the general public whose data was being shared. For data repositories, this included having established protocols in place for sharing sensitive data and vetting data accessors to

“...you will never eliminate risk, and you must never get bogged down into what they call “Frankenstein scenarios”, which is we did this, this and this. They talk about having process, being able to demonstrate due diligence, being able to say these are the processes we have in place; we followed the processes, and demonstrate you can follow the process. That’s as good as you’re going to get.”

-Interview Participant

ensure they were using data appropriately. The majority of other data repositories interviewed had strict accessing protocols that included screening potential accessors, ethics approval, only making de-identified data available that passed confidentiality screens, and providing access through secure VPN channels. Different requirements for accessing data will be discussed further below.

Importantly, the document review indicated that it was the data holder's responsibility to oversee data protection (Scottish Health Informatics Programme, 2010). To address data protection SAGE's Operational Advisory Committee sought counsel from legal, ethical, and privacy experts and implemented policies and procedures that aligned with consultations.

<i>Key Considerations</i>	<i>Related Findings</i>
Implement data accessing protocols and vet potential data accessors	Risk can be effectively managed using a variety of techniques

Broader Culture of Data Sharing and Repositories

Cultural changes to data sharing are occurring, including recognized benefits and increased education; however, some barriers still exist. This section explores what cultural changes have taken place, barriers experienced by researchers, institutions, and data repositories, and the concrete ways in which change is being supported or barriers addressed. It is important to note that in this and the following sections the research pillar of data sharing is emphasized. Though the community, and to a certain extent, administrative pillars are represented, many of the perspectives are shared by researchers in those respective fields, further contributing to an emphasis on the research pillar. This reflects an area for further consideration.

Culture Change

According to interviews and the literature review, the field of data sharing and repositories is evolving and there is growing recognition of data sharing benefits (Research Data Canada, 2011), such as preservation of data assets, and increased knowledge and collaboration in the scientific community (Medical Research Council, 2016). Researchers, agencies, and institutions are becoming more data driven, meaning support for data sharing will likely grow and data management will become more heavily relied upon. Funders and governing research bodies are increasingly offering funding to those that use secondary data or by requiring funded researchers to share their data. For example, the National Institutes for Health (2003) expects researchers receiving \$500,000 or more in funding to share their data. Further, funding proposals to the Economic and Research Council (2013) require researchers to demonstrate no other data is available for secondary use. This incorporates data sharing into research proposals and design. Interview participants said community organizations are also relying on data to inform practice and to conduct assessments or evaluations.

Interview participants emphasized the benefits of data sharing, including its ability to:

- Advance science and reduce duplication - Existing data allows data to be used in multiple ways to answer new research questions.
- Increase efficiency and productivity - Reuse of data minimizes the need to collect new data, which is often time consuming and costly.
- Make the most of funding dollars - Data sharing and reuse of data means that funding can be put towards analysis rather than collecting data that may already exist. Additionally, sharing of publicly funded data ensures public engagement and maximizes data use (Research Data Canada, 2011).

For the evolution of data sharing to endure, interview participants and the literature emphasize ongoing education and promotion. Education includes training new researchers and providing support to universities in regards to secondary data use, its approaches and methods, and how to access it (Lee, Sung, Barnett, & Norris, 2016). One interview participant stated that a multi-directional approach (i.e., ground up and top down) that provides different messages to data users, funding bodies, institutions, and community organizations is beneficial. Promoting data sharing will also encourage improved data preservation methods and analytic tools, increase the number of skilled workers to manage the data, and support Canada in becoming a leader in research (Research Data Canada, 2011).

Facilitating education is an important area of focus for SAGE. Educating others on the benefits of data sharing and approaches to facilitating learning was mentioned in the majority of documents reviewed. Based on interviews with the SAGE team, an example of supporting education is the grant that allowed researchers the opportunity to work with secondary data. Other educational opportunities included training, webinars, and lunch and learns. Additionally, data sharing was being promoted to universities, government, community agencies, researchers, and the general public.

<i>Key Considerations</i>	<i>Related Findings</i>
Work with PolicyWise and SAGE grant recipients to encourage data sharing	Funders are recognizing the benefit of data sharing and many require funded researchers to share their data (i.e., NIH, ERC)
Continue promoting data sharing through grants, training and conferences while considering areas for expansion	Education and promotion targeted at data users, funding bodies, community organizations, and institutions supports cultural change related to data sharing
Establish champions of SAGE that will promote the initiative, and educate others through mainstream channels and social media	

Cultural Barriers

Cultural shifts in data sharing are occurring; however, a number of barriers still exist. This section discusses cultural barriers to data sharing, including reluctance of researchers to share data, lack of knowledge and understanding of data repositories and secondary data use, costs of producing, accessing or depositing, and housing data, and funding constraints.

Across interviews and the literature review a common theme was reluctance among data producers to share their data (Research Data Canada, 2011). According to interview participants, researchers are often reluctant to share their data due to concern their data will be used without proper credit or in ways it was not intended. These concerns sometimes resulted in a 'mine' versus 'yours' mentality or misunderstanding intellectual property rights.

Limited knowledge or understanding of data repositories was commonly discussed by interview participants. Interviews revealed that many researchers are unaware data repositories exist or how they work. As a result, researchers often default to collecting primary data rather than using secondary data. Lack of awareness contributes to researchers not considering depositing data or the steps necessary to do so, such as obtaining consent by participants to share data. For data accessors, there can be confusion about which repositories are appropriate for their data use needs and area of focus.

According to internal and external interviews there are a number of misconceptions around collecting data. First, there is a perception that collecting primary data is superior to secondary data because it is considered original and can be tailored to the specific focus of the researcher. Though there is a growing trend towards secondary use of data the benefits are still not fully recognized. One participant noted that there is currently no method that supports secondary data use or a combination of primary and secondary data analysis. Misconceptions are further reinforced by funders that mainly provide grants and other funding to researchers collecting and analyzing primary data. Additionally, and according to interviews, the research field is moving towards open access platforms, which provides greater access, but the data is often not reviewed, resulting in data that is missing information or cannot be replicated and contributing to the perception that secondary data is not as useful.

Second, many researchers are under the impression that in order to use secondary data they must collaborate with the original researcher. While every data repository has different processes and regulations around secondary data use, many do not require researchers accessing secondary data to collaborate with the data producer. SAGE addresses this concern by allowing principle investigators to decide what level of involvement they would like to have with others accessing their data ranging from minimal or no involvement to full collaboration and co-authorship.

Funding requirements and constraints are another barrier for data accessors, depositors, and repositories. Interview participants noted that although analyzing secondary data is often cheaper than

collecting and analyzing primary data, there can still be significant costs to researchers accessing secondary data, and which funders are less likely to offer financial support.

There are costs to researchers collecting data to be deposited. For example, several interview participants discussed repositories' desire for large datasets and emphasized the cost of developing a research team with the capability of collecting large amounts of data sufficient for deposit. Researchers just entering their field often experience greater challenges collecting large amounts of data. Interview participants also discussed costs associated with creating appropriate metadata labels to ensure other researchers can use the data. One participant referred to this as developing "independently understandable" data, meaning that each variable is described in depth including how it was produced and measured. Though funding is more readily available for collecting primary data, funding is often competitive and not available for the duration of the research. Given cost and time there is less incentive for researchers to collect data for deposit or to share it with other researchers.

Data repositories also experience challenges or barriers to funding according to interview participants. There are costs associated with curating data and the infrastructure required to house data securely. Funders may be hesitant to provide financial support and infrastructure when they are unsure about the longevity of the repository, especially if it is not linked to an institution such as a university. Obtaining financial support requires ongoing relationship building and partnerships.

On a systems level, an OECD (2007) report based on consultations with experts, research institutions, and policymakers found that international frameworks for data sharing were absent in many of the OECD member countries. The authors suggest that research policies and guidelines need to be adapted to support advances in the use of information and communication technology in research, knowledge sharing, and collaboration. Additionally, Canada is one of the only developed countries that do not have a data management plan for publicly funded research and data (Research Data Canada, 2011). This results in challenges with long-term preservation and development of analytic tools. It also contributes to researchers using incongruent data due to unavailability of the necessary data. These cultural barriers can and are being addressed.

Supporting Change and Addressing Barriers

There are a number of ways that SAGE, researchers, community organizations, funding bodies, and data repositories are supporting cultural change and addressing barriers related to data sharing. As one SAGE team member stated, challenges and barriers are actually opportunities to learn and grow.

Across interviews and the document review a common theme was educating researchers, the public, and institutions about the benefits of data sharing and secondary analysis (ACCFR, 2013; ACCFR, 2016). On a smaller scale, several interview participants discussed how they demonstrated the benefits of

"Let's train a generation of researchers! Research dollars are shrinking and the competition for grants is brutal...Why do you have to go that way? The data is collected. There is so much you could do with that data that is already available."

-Interview Participant

data sharing and secondary data use to researchers, including how their data could be linked with administrative data to make it more powerful, or how other researchers could use their data to answer different research questions. SAGE and a number of other interview participants also had champions for data repositories that promoted data sharing and secondary analysis. Additionally, public channels such as Twitter and Facebook were used for education, advocacy, and promotion.

At a systems-level, researchers and data repositories, including SAGE, were working with each other, and in some instances their respective provincial governments, to create dialogue around shared concerns or challenges, legislation, and promotion of data sharing (ACCFCR, 2016). Many of the concerns were related to privacy and protecting participant data. One solution was to develop shared language on ethics and consent forms. Though this is not standard across Canada yet, many organizations and data repositories were developing processes to streamline ethics, consent, and data sharing agreements with the hope of establishing similar processes provincially and nationally. Others were speaking to governing bodies about research participants willingness to have their data shared granted they are not identifiable and their data is being used for research purposes. The Research Data Strategy Working Group hosted the 2011 Canadian Research Data Summit to address a number of these barriers, including the absence of a national data management plan for publicly funded research. The Summit resulted in a draft national strategy, which indicates a move towards policies that promote data sharing in Canada. Additionally, there were concerns related to funding. Many funders are beginning to include data sharing into their conditions as a way to maximize funding dollars and promote data sharing. One data repository addressed funding concerns by working with their provincial government to have their budget included in the provincial budget.

According to interview participants, providing opportunities for research students to learn about secondary data use and sharing was another means of contributing to culture change. In one instance, students were granted an award for accessing existing data to cover the access fees that researchers normally pay. The purpose was to promote secondary data use and provide access to data that students might not be able to afford to access otherwise. Similarly, SAGE offered a grant to researchers as an incentive to use secondary data. Another interview participant discussed courses and workshops their organization offered to students, including quantitative methods and stats course, as well as training on how to use their data, and data management and curation. Other entities were promoting methods that combined new and existing data.

Interview participants also described the ways in which researchers were being supported throughout their careers in regards to data use and sharing. Many interview participants were working with researchers to assist them with thinking about and designing their projects in the early phases. This encouraged researchers to consider the possibility of accessing existing data and ensuring that consent forms accounted for the possibility of data depositing when the research was complete. Likewise, SAGE raises awareness about data sharing and encourages researchers to think about secondary data use and deposit before they initiate their research rather than it being an afterthought. One interview participant described the importance of preserving data even after researchers retire. Therefore, they

used grants and other funding to ensure the longevity of data. Additionally, due to the changing landscape of data sharing in Canada, SAGE is constantly repositioning itself to address the needs of data users and producers.

<i>Key Considerations</i>	<i>Related Findings</i>
Demonstrate the usability of secondary data by advertising projects completed using existing data	There is a perception that primary data is superior, or researchers are unaware secondary data is available to them
Promote methods for secondary analysis among students and university faculty	
Support long-term preservation for ongoing data management cycles and storage of data for researchers once they leave the field	
Encourage funders to provide more funding opportunities for researchers collecting data for deposit or conducting secondary analysis	The cost of producing data with large enough samples is high, and though secondary analysis is cheaper, it is still costly and time consuming
Collaborate with other data repositories and governments to promote and facilitate the development of frameworks or policies that support data collection, preservation, and sharing	There is a lack of international ethics frameworks, which hinders international data collection and data sharing. Additionally, Canada does not have a data management plan for data collected using public funds.
Develop standard ethics and consent language around data sharing	
Involve research participants in larger discussions about consent and data sharing	

Goals, Outcomes, and Impacts about Data Sharing

This section discusses goals, and intended or achieved outcomes and impacts of data sharing or data repositories. These include prioritizing data sharing processes, building relationships, strengthening and mobilizing capacity, and increasing impactful work.

Prioritizing Data Sharing Processes

Across all data components the importance of engaging researchers early in the research process was highlighted. Working with researchers early helps address different concerns or considerations at each stage beginning with project planning (Humphrey, 2006). According to the literature, having a quality research plan that includes data sharing mitigates ethical concerns related to secondary analysis (Economic and Social Research Council, 2013) and accounts for how roles and responsibilities change throughout the research lifecycle as well as the data lifecycle (Corti, Van den Eynden, Bishop, & Woollard, 2014). These plans should be included starting with funding proposals and throughout the

research process. Based on interviews with participants and the document review, contributing to and helping develop standards for data sharing agreements was a goal of SAGE's and other data repositories. Dove et al. (2016) states that clarity in data sharing agreements and ethics processes need to be achieved. Specifically, once an overarching ethics review process has been agreed upon it should be equally applied to all projects with similar risks and in all jurisdictions while allowing for local accommodation. Interview participants discussed their goal of developing a common data sharing governance framework, including ethics processes that could be universally accepted and tailored for each province.

Other goals for data sharing processes as described by interview participants included developing better ways of storing and preserving data, contributing to the creation and adaptation of leading practices, and having a broad impact on data sharing.

<i>Key Considerations</i>	<i>Related Findings</i>
Support researchers early in the research process starting with funding proposals, if possible, to leave open the opportunity for data sharing	Clear research plans that include potential data sharing mitigate ethics and consent concerns and account for changes in roles or responsibilities throughout the research/data lifecycle
SAGE is engaged in several leading practices that others would benefit learning from. Continue participating in opportunities to share SAGE's approach and learn from others	Overall goals of data repositories included developing better storage and preservation approaches, and contributing to leading practices on data sharing
Publish articles, reports, and other documents	

Building Relationships

Building relationships was emphasized across interviews, the document review, and to a lesser extent the literature. As Humphreys (2006) notes, without partnerships long-term preservation and access to data would not be possible; however, in order to build partnerships we must understand the research process and areas or gaps that can be addressed through data preservation and access.

Three types of relationships were discussed by interview participants: other data repositories, data producers and accessors, and policymakers or institutions. Each relationship served a different function. Relationships with other data repositories, or those doing similar work, were viewed as learning opportunities and potential collaborators or partnerships. For example, several data repositories mentioned their connection to the Population Data Linkage Network or Strategy for Patient-Oriented Research (SPOR). Similarly, SAGE was involved with Research Data Canada to coordinate ideas around data repositories with the goal of being visible nationally. SAGE had also identified other strategic connections, such as the ethics Tri-Council, SPOR, and Alberta Innovates to name a few.

Relationships with data producers and accessors were fundamental according to interview participants. Developing strong and trustworthy relationships with researchers was needed before repositories could inquire about data. This required extra time and effort while demonstrating the benefits of sharing data. One data repository built relationships with potential data producers by completing research projects for them that linked their data with data already held in the repository. The goal was to show what the producer's data was capable of when linked with other data or used to answer different research questions. They also worked with data producers to determine restrictions for researchers accessing their data. Several data producers requested strict regulations around who could access their data, which they would often relax after the repository demonstrated their ability to protect the data and privacy of their research participants. The document review showed that developing strong partnerships between SAGE and principal investigators was a priority, and emphasized the importance of addressing principal investigator concerns related to data ownership and protection of privacy (Adair, 2012; ZGM, 2016).

There were also a number of benefits to developing relationships with data accessors. Interview participants, both internal and external to SAGE, noted that SAGE served as a connector between data producers and accessors that could promote collaboration. For example, connections between researchers that had deposited data into SAGE and students looking for access to secondary data related to children and families could be facilitated. Facilitating these connections was described as a driver for data producers and accessors to work with SAGE. Additionally, PolicyWise and SAGE grant competitions contributed to data producers being able to conduct research and the likelihood they would share data. One goal of SAGE mentioned in the document review was facilitating "super users" that could contribute to dialogue around data sharing and promote SAGE (Adair, 2012).

Other relationships discussed by interview participants and in the document review were those with policymakers and institutions such as universities. Across the data repositories interviewed, developing relationships with policymakers to answer policy relevant research questions was a common goal and desired outcome. Developing these relationships also created opportunity to work with policymakers and government officials on data sharing legislation. As for other relationships such as those between university researchers and data accessors, The Scottish Health Informatics Programme (SHIP) Guiding Principles and Best Practices (2010) document states that cross-sectoral data sharing is beneficial. These relationships open the opportunity to create reciprocal privacy standards and MOU's. Additionally, interview participants stated that SAGE can serve as a representative for universities and researchers (i.e., privacy and ethics legislation) because of their connection to government entities. Each of these facilitates cross-sectoral collaborations and networks.

The CDCA Marketing and Communication Plan recommended that SAGE leverages relationships already formed through PolicyWise to begin building its reputation (ZGM, 2016). Once SAGE is well established it can begin building a more individual reputation.

<i>Key Considerations</i>	<i>Related Findings</i>
Continue engaging and building relationships with universities, community agencies, and potential data accessors or depositors	SAGE has developed many partnerships and collaborations, particularly in the area of government ministries and ethics bodies
Facilitate collaborations or partnerships between students and junior faculty and researchers interested in depositing their data for secondary analysis	
Work with cross-sectoral partners to develop privacy standards, MOU's, etc. to facilitate cross-sectoral data sharing	Cross-sectoral data sharing and partnerships are beneficial and lead to a more holistic understanding
Represent the needs of other sectors through existing relationships with government ministries	
Continue leveraging relationships developed by PolicyWise	

Strengthen and Mobilize Capacity

Strengthening and mobilizing capacity was not a theme discussed in the literature; however, it was a goal of most interview participants. Due to the wide range of interview participants building capacity occurred through several different avenues such as research production, data sharing, answering new research questions using existing data, or by preserving and linking data. The overall goal was to make data available and accessible. One repository addressed concerns related to long-term data preservation and availability by offering preservation packages to researchers that did not have the resources to convert their archival data.

More specific goals of capacity building for SAGE were developing students' potential, which was further supported through the SAGE grant. Facilitating secondary data analysis among students will foster highly skilled employees and address current gaps in managing research data in Canada (Research Data Canada, 2011). Further, SAGE's goal to link across administrative, research, and community data encourages all researchers to contribute to or access data regardless of whether they are connected to an institution.

"We are trying to get all this data in one place so eventually we can link all the data together and answer more complex questions to support evidence-informed policy and practice. That's our ultimate goal, but also what we really want to accomplish with SAGE is to build capacity within research and within evidence-based policy. We try to promote that collaboration between people who may not necessarily be collaborating without this connection."

-Interview Participant

<i>Key Considerations</i>	<i>Related Findings</i>
Consider offering preservation packages as an added service for researchers without the tools necessary	The overall goal of data sharing is to increase access and use of existing data
Using SAGE's expertise to offer courses and training on data management and curation to students, junior faculty, or other researchers	There are few skilled workers in the area of data management and curation in Canada

Increasing Impactful Work

Across all research components the goal of sharing data was to increase the value of data and to advance science. A goal of interview participants was producing data that represented services actually used by citizens.

Much of the literature on data sharing is focused on health; however, the sentiment that data sharing can significantly increase scientific knowledge and translation can be applied to various fields of research. According to the OECD (2007), international frameworks for data sharing require further improvement; however, global scientific databases are growing, increasing data sharing and access which:

- Reinforces open scientific inquiry;
- Encourages diversity of analysis and opinion;
- Promotes new research;
- Makes possible the testing of new or alternative hypotheses and methods of analysis;
- Supports studies on data collection methods and measurement;
- Facilitates the education of new researchers;
- Enables the exploration of topics not envisioned by the initial investigators; and
- Permits the creation of new data sets when data from multiple sources are combined (p.10)

Interviews with SAGE and the document review highlighted SAGE's goal to facilitate research productivity, including better research conducted in shorter periods of time, new scientific discovery, rapid creation of knowledge, and archived data that could be used again. This could be achieved through access to existing data because it minimizes the time and cost of collecting primary data and allows data to be used in new and creative ways. Further, SAGE has goals to push the boundaries of innovation, facilitate cross-discipline research, and be a leader in Canada and globally as the first data repository focused on children (Adair, 2012; The CDCA Evidence-Informed Policy and Practice through Optimal use of Research and Administrative Data, 2015).

Interviews with external models and stakeholders revealed that data sharing has the possibility of producing data about the services accessed by people. They suggested that other approaches, such as randomized controlled trials, have many benefits but the results often do not represent the people actually accessing services. Therefore, access and the ability to link across different types of data, such as community and administrative data would result in more representative samples and findings that

could be used to increase or improve services in the service areas being accessed most. This supports SAGE’s goal of being involved in projects aligned with PolicyWise’ mission “to develop and integrate evidence to inform, identify and promote effective public policy and service delivery to improve the well-being of children, families and communities in Alberta, Canada and internationally” (Strategic Business Plan, 2014). Additionally, SAGE intends to include administrative data, which would allow for more complex and holistic analysis that could be used to support improved policy development.

SAGE interview participants stated that through the learning process SAGE has further contributed to increased productivity, impactful work, and the field of secondary data use by sharing their findings in white papers, publications and presentations.

<i>Key Considerations</i>	<i>Related Findings</i>
Continue to work with government ministries to merge or incorporate data into SAGE	Linking data offers an accurate picture of the services needed or being accessed by people, which can be used to improve policy decisions

Operations

This section discusses the operational aspects of external models, stakeholders and SAGE. Therefore, only themes from interviews and the document review are presented. Operational themes include promotion and visibility, requirements of the data, requirements for accessing data, providing support to data producers and accessors, remaining up to date on leading practices in data sharing and repositories, and sustainability.

Promotion and Visibility

Interviews with SAGE and external models or stakeholders emphasized the importance of promotion and visibility. Promotion directly tied into visibility as it was about being recognized by people not only for data, but for involvement in the research community. According to interviews with SAGE increased awareness contributes to potential future collaborations and ambassadors of SAGE. Ambassadors were described as a key to success because they could contribute to increased depositors and accessors as well as the visibility of SAGE; however, SAGE interview participants felt that more needed to be done to get the word out in order for SAGE to be viewed as “the place to go for data”. Lee, Sung, Barnett & Norris (2016), state that strategies for promotion should change with the needs of users and in response to new technology. The best approach for increasing SAGE’s visibility is still being explored.

Interview participants had several different approaches for promotion and achieving visibility. Some were using a broad networking approach for promotion. For example, the majority of external models and stakeholders interviewed attended some form of conference, presentation, or event to promote their work. Some used social media like Twitter or Facebook to post information or messages, though a number felt that social media was not as effective as they had anticipated. One data repository published an article in an international journal outlining the process for accessing their data.

Another technique for promotion and visibility was using a targeted approach. Several interview participants engaged universities, different faculties, or students to demonstrate how data could be used for secondary analysis and to present opportunities for data deposit. This was particularly helpful for students that could not collect primary data. The majority of interview participants also engaged in strategic networking opportunities and person-to-person conversations about potential collaborations, partnerships, or data deposit. One data repository had a designated staff member that solicited collections from researchers. More specific examples included offering memberships for collections, creating user stories that attach easily readable information to every collection in the repository, or offering student grants.

SAGE used a number of similar approaches. Based on interviews, SAGE participated in conferences, presentations, and lunch and learns. SAGE also connected to and affiliated with universities and community agencies, or reached out to researchers and past grant recipients (PolicyWise and CIHR) regarding data sharing. Additionally, SAGE held a grant competition as incentive for researchers to conduct secondary analysis.

SAGE also used a number of approaches that were not specifically discussed by other models or external stakeholders. For example, SAGE fostered relationships with researchers that became super users and champions. SAGE has also worked closely with different Ministries, the OIPC, and ethics boards to keep up to date and facilitate changes to processes and help develop data sharing standards. These approaches have made SAGE visible in the research community.

<i>Key Considerations</i>	<i>Related Findings</i>
SAGE is continuing to explore the best approach to promotion and visibility. A combination of broad approaches (i.e., conferences, social media, publications) and targeted approaches (i.e., grants, reaching out to previous researchers) is beneficial	Adapt promotion based on the changing needs of users and technology

Data Requirements

Data requirements refer to areas of focus, type of data, or how the data is collected, scrubbed, or formatted prior to data repositories accepting it for deposit or a researcher using it for analysis. This section also discusses considerations for general versus specific data repositories.

Based on interviews with external models and stakeholders, important elements for deposit include demonstrated usability, high quality², and well described or independently understandable data. One interview participant said data repositories can accept data that isn't well described, but not all resources should go towards preparing it.

² High quality data was described as the collection was complete, no information was missing, and the labels were correct.

The majority of data repositories interviewed required all data to be de-identified, even if they were a prescribed entity entitled to hold identifiable administrative data. One such entity required data to be completely de-identified and coded numerically before accepting any data. Data also had to be migrated to non-proprietary formats in order to make it more understandable and address ownership concerns for long-term preservation. Another repository accepted identified data but restricted who had access within the organization. Those with the highest access clearance would replace identifiable information with a numerical code before providing access to others within the organization. The data was often subjected to confidentiality reviews to ensure no one could be identified. Additionally, most required incoming data to have consent from research participants to be shared, with the exception of administrative data, which had different regulations in each province.

“...the more you can do to ensure the quality of your collection is high, the better position you are in to play a real difference for your community...curation is still very important and the more targeted effort you can put into your collections the better. Your community will thank you and your reputation will shine and stand out.”

-Interview Participant

In regards to the type of collections offered by data repositories, a number of participants stated there were different data considerations. For generalized curation a robust collection that researchers with different interests and topic areas can access was suggested. Specific curation on the other hand should be tailored and high quality.

SAGE also had a number of data requirements. For individual level data, SAGE looked for data that aligned with the social sector, was a relevant topic area, and cross-sectoral. SAGE’s niche is in the area of child and family wellbeing. Large data sets were considered ideal, but there was recognized value in smaller data sets as well. Involvement with researchers during the early stages of data collection helped ensure metadata was high quality. Further, SAGE has not turned away any data sets during the initial phases.

<i>Key Considerations</i>	<i>Related Findings</i>
Accepting de-identified data may ease privacy concerns associated with administrative data	The majority of data repositories holding administrative data only accepted de-identified data or restricted access within their organization even when it was not required by the data partner
SAGE currently has a focus on social sector data related to children and families; the scope of SAGE can be narrowed, maintained or expanded based on the needs of data depositors and accessors	Data repositories usually offer either general data collections or specific data collects, each of which have pros and cons

Data Use Requirements

Requirements for data use are the criteria researchers must meet before they are granted access to data from a data repository. The criteria differs by repository, with some being open access (i.e., publicly available) and others being restricted.

According to interviews with external models, open access data repositories generally did not have criteria researchers needed to meet for access. Metadata was made available on the website where it could be viewed or downloaded. Some data repositories were restricted access, but did release publicly available scrubbed versions of datasets.

Other data repositories had stricter requirements researchers had to meet before they could access the data. In all cases an application form or something similar was necessary. Most did not have set times for when applications could be submitted, but some smaller repositories were considering set times due to the increased number of requests and limited resources to keep up. Applications allowed data repositories to determine what the data was going to be used for, specific criteria the researcher needed to meet, and if any special processes were necessary. One repository required accessors to go through a feasibility process to ensure that what they wanted to do was feasible and the data was available. According to Lee, Sung, Barnett, & Norris (2016), feasibility tests improve the likelihood of the project being successful and the data being useful. Interestingly, all interview participants said they did not judge the quality of the research in the application because they felt it was the position of the research community during the peer review process.

The majority of data repositories, but not all, required ethics to be obtained prior to data being provided. In some instances, a review process determined if individual consent was also necessary to use the data; however, this may also have been accounted for during the depositing phase if consent was required by the repository for research participants to have their data shared. One repository that offered biological samples further required researchers to go through a peer review process prior to access because biological samples are a depletable resource that cannot be easily replaced.

Once applications were accepted, each repository had different processes for sharing the data. According to one repository, researchers would be given remote access to some data through a secure VPN “tunneled” into the organization. This was due to strict regulations by data partners about where and how data could be stored. In another case the repository did not allow internal researchers direct access to data. Rather, the scientist would pose a research question and an internal analyst would run the data. Other repositories completed risk identification processes to ensure the likelihood of individual identification was limited and cells with less than five cases were not reported. This approach aligns with Statistics Canada’s requirement for data sharing.

“One of our interesting things over the last little while has been to change even the mindset of our staff towards being much more customer focused and flexible, and this flexibility is critical. There are general principles in which you operate, but there is a lot of wiggle room and flexibility, and for our staff to have the confidence to say to the researchers, ‘we don’t have that, but have you considered this?’”

-Interview participant

According to interviews with SAGE, they also have a number of processes in place for data accessors. SAGE required data accessors to complete a request, which was reviewed for merit, feasibility, and appropriateness. The review panel includes someone internal to SAGE or the CYDL, the original data producer, and an external reviewer. The review not only determines suitability, but also the skill and budget of the researcher requesting access to the data. Upon approval of the request, a legal secondary data use agreement is completed, which includes acknowledging the data source (SAGE and original data producer). The accessor is then provided a customized dataset and access to analytic tools such as SPSS or SAS. Access is typically through secure remote access; however, access through a physical enclave room or secure file transfer protocol is also possible.

<i>Key Considerations</i>	<i>Related Findings</i>
Data repositories holding administrative data also had policies related to internal staff access; since holding administrative data is a goal for SAGE, consideration for who within SAGE/PolicyWise will have access may be useful	Most data repositories, including SAGE, have similar data access processes, including requests or applications, reviews, and data sharing agreements; there was variation in how data is shared

Other Operations

This section discusses other operational considerations of data repositories mentioned by interview participants, including the amount of support they provide to data producers and accessors, how they remain up to date on leading practices, and sustainability of data sharing and repositories.

Providing Support and Flexibility

Of the external models and stakeholders interviewed most offered some form of support to data producers and accessors; however, this varied widely. Several offered support by uploading information and access to forms online. Online forms offered examples and guidance in regards to how they should be completed. Telephone numbers were usually available as well for applicants to inquire with questions or challenges. A number of data repositories offered more extensive support, which was often referred to as “handholding”. Repositories recognized that people accessing data have different levels of experience and support needs. Therefore, most were flexible in their approach to providing support. These repositories also offered metadata and data dictionaries that described the data in-depth.

Across interviews with SAGE and the document review, providing support to data depositors and accessors was emphasized. Offering support was described as a “value-added” service (ACCFR, 2013); however, the documents discussed the importance of not competing with other services, and instead collaborating, referring out, or leveraging existing technical and business infrastructure where needed.

According to SAGE interviews, support is provided at all stages of the research process. Engaging with researchers as early as possible ensures ethics and consent are in place for later data sharing. The process is hands on to make it easier for accessors to conduct secondary analysis and understand the

“how tos” of data. SAGE interview participants said community organizations need more “handholding” and education than others in regards to data and data sharing because collection and analysis in community services has few established processes yet. Other supports offered included a range of metadata, ethics and legal support, and fostering collaboration between researchers and others.

SAGE was also flexible in their approach to services offered. The relatively small size of SAGE allows for more tailored services and supports as well as timely responses to researchers, which larger data repositories may not have the flexibility to do. According to SAGE interview participants, flexibility contributes to stakeholder buy-in. SAGE has been able learn from stakeholders in order to offer the services most needed. Additionally, flexibility allows data depositors to choose their level of involvement with data accessors. Nonetheless, SAGE cannot suit everyone’s needs, which was described as part of the ongoing learning process.

Staying up to Date on Leading Practices

Common approaches to staying up to date on leading practices by external models and stakeholders included belonging to mailing lists, attending conferences, communicating and networking with peers in the field, reading publications, and accessing training or professional development opportunities. A number of interview participants stated that their organization held meetings to keep each other informed of important developments related to their positions. One interview participants described being “pushed along” by innovative researchers in their organization. Remaining innovative was challenging in some instances because of the infrastructure and funding required. Another challenge discussed by interview participants was findings answers to questions related to changes in the field. Occasionally there was confusion about where to go, and when an answer was finally acquired something would change again.

For SAGE, remaining involved through connections with others, such as the OIPC, Alberta Innovates or researchers, was described as a way to remain up to date. Importantly, interview participants said SAGE tried to stay up to date on the broader research world in order to understand leading edge practices. Additionally, SAGE has an implementation group that meets regularly.

Sustainability

Sustainability referred to a number of things including techniques for data preservation, cost recovery mechanisms, and maintaining relevance and presence as an organization.

A major goal of the external models interviewed was ensuring long-term preservation of data, which was referred to as an asset. Not only did this include migrating data into more accessible and readable formats, but also implementing strong security systems to protect the data.

“But the success of SAGE will ultimately be successful because we are ambassadors of the product...the success of SAGE will depend on our ability to hold the long view and respond to the needs of our researchers and research custodians.”

-Interview Participant

Several positive changes related to data preservation and storage were listed by interview participants, such as the ease at which repositories can be started, cheaper storage and computational resources, and better access to the internet. Nonetheless, the human cost of maintaining data repositories was still high according to interview participants. Many of the repositories costs for data access were related to maintaining infrastructure and security, providing support, and preparing data for secondary use. Therefore, several repositories had or were exploring cost recovery mechanisms. For example, one repository offered memberships for certain data collections, while another had precisely calculated the cost of preparing data for access and charged fees to cover those costs. The rest of their funding was through grants and contracts, which ebbed and flowed.

Other sustainability considerations raised by interview participants were related to maintaining their presence as an organization. That meant working with data partners at a level they were comfortable with, continuing to develop relationships, and demonstrating the relevance of their work. One interview participant emphasized the need to reassure users that the repository will be around over time, but also developing a contingency plan that includes partnering with another organization in the event the repository ceases to exist.

According to SAGE interviews, the key to success and sustainability was leadership, communication, developing relationships, and attention to detail. It was for this reason they were working with other initiatives and learning from others in order to keep the momentum. The established reputation of SAGE and PolicyWise was described as a benefit moving forward.

<i>Key Considerations</i>	<i>Related Findings</i>
As SAGE grows the level and type of support or flexibility offered may need to be adapted	The majority of repositories offered some support, recognizing depositors and accessors have different levels of experience and support needs; however, experimentation with the level of support and resources put towards it was common
Some repositories had or were considering cost recovery mechanisms to supplement funding	Maintaining repositories has associated costs
A number of data repositories had agreements with other organizations or repositories to hold the data in the event the repository ceased to exist	Building trust with data repository users includes reassurance that it will be around over time, but having a contingency plan in case

Governance

This section begins by describing the governance structures of other data repositories based on external interviews. Following, there is a discussion of SAGE's governance structure and Advisory Committees, consisting of the document review, and interviews with SAGE and external stakeholders.

The majority of other data repositories had a committee or council, which varied in size and representatives; however, interview participants stated that committee members should be carefully

considered, and decisions should be made regarding whether committee representation should be research or process heavy. Interview participants also recommended including ethicists and legal representatives, something SAGE did early on.

Unrelated to individual repositories, interview participants thought that starting a larger council for data repositories in Canada that addresses barriers and challenges commonly experienced would be beneficial. For example, the UK Biobank created an arms-length governance council to provide options and incidental findings. Implementing a similar model could assist with developing a consistent legal framework in Canada.

Advisory Committees

The Advisory Committees (ACs) were specific to SAGE; however information from internal and external stakeholders, as well as the document review were incorporated.

During the initial stages SAGE had a working group that eventually branched into an Operational Committee and a Strategic Committee, each with their own focus. The Operational Advisory Committee (OAC) focuses on the day-to-day details, legal and ethical aspects, leading practices, procedures, improvement, and IT, but does not report to the board. The Strategic Advisory Committee (SAC) focuses on strategic direction, increasing the profile and partnerships of SAGE, funding, education, further opportunities, and overall issues, and does reports to the board. Additionally, the ACs provide direction and expertise on ethics and privacy. Interestingly, none of the interview participants mentioned the SAC reporting to the board, indicating they may not be aware of this. Mentioned in the document review, having a committee involved in governance is supported by evidence (Scottish Health Informatics Programme, 2010). Though there is a distinction between the OAC and the SAC, the document review revealed blurring in regards to what is discussed in each of the Committee meetings.

The ACs are a form of engagement and consist of relevant stakeholders SAGE can share ideas with and seek advice from (ACCFRC, 2013; ACCFCR, 2016). For example, discussions about the scope of SAGE (i.e., referrals, maintaining focus on children) have been brought up in both committees (CDCA Operational Advisory Committee Meeting, 2015). Additionally, quality improvement has been incorporated into SAGE and consideration given to tracking citations (CDCA Operational Advisory Committee Meeting, 2015). Citation tracking has not occurred to date because SAGE is working with their first batch of users. At one time general public members were included as stakeholders in the OACs Terms of Reference, but that has since been removed, which aligns with other models interviewed that do not include public representation. Further, the document review highlighted the need to incorporate Indigenous representation into the ACs.

A common theme across all interviews and the document review was the infrequency of meetings for both the OAC and SAC. According to interview participants, the OAC meets approximately once a year, which has made ongoing communication challenging. The documents suggest the OAC meets three times per year, and the SAC meets twice per year; however, meetings have not occurred as frequently

as outline in the Terms of Reference. Additionally, the document review revealed that items discussed at previous meetings may not be followed-up on or are deferred. Long periods of time between meetings may contribute to this.

Interviews with external stakeholders suggested increasing the frequency of meetings and communications, particularly between OAC committee members. One solution offered was providing some form of update to OAC members between meetings with detailed information about SAGE. Others suggested meeting in person occasionally, rather than always by phone. Additionally, some external interview participants felt there was high turnover in the ACs, which was not perceived negatively, but was thought to be slowing SAGE's progress due to changes in direction.

<i>Key Considerations</i>	<i>Related Findings</i>
A larger council or committee to represent the interests of Canadian data repositories was suggested	There is limited representation of data repositories on a national or international level, contributing to inconsistent policies and legislation
Implementing two committees places SAGE ahead of other data repositories	Developing a governance committee or council was good practice
Increasing the frequency of meetings, sending out regular updates, and meeting in-person occasionally may improve communication and the overall experience of AC members	The OAC and SAC were not meeting as frequently as anticipated

Opportunities and Future Plans

This section describes the different opportunities and future plans described by interview participants, including what is working well, areas for improvement, and considerations or potential next steps.

What is Working Well

Across all interview participants there was a consistent belief that most partnerships were working well. Partnerships benefited both researchers and data repositories by offering funding, supporting data collection, analysis, and preservation, and encouraging broader dialogue related to data sharing. Additionally, partnerships bridged some of the gaps between researchers, community agencies, and institutions such as universities. These were viewed as facilitators of data sharing.

Interviews with other data repositories indicated that most felt their data accessing and depositing processes were working well. The majority of data repositories had also been established for a long period of time, providing them the opportunity to work through different issues or considerations.

Specific to SAGE, external models and stakeholder interview participants stated that the processes related to privacy and ethics were well thought out and implemented. The data itself was also highlighted as an aspect that was working well. Interview participants said the data was clean and

prepared for use, with the capability of supporting different issues and program or policy development. Further, the support offered by SAGE to data accessors and depositors was viewed as invaluable.

Areas for Improvement

As discussed above, areas for improvement emphasized by all interview participants were the overall cultural barriers, including lack of knowledge about data sharing and data repositories, ethical and privacy concerns, and ownership of data. Interview participants felt themselves and others could improve by offering more education to researchers, institutions, and the public about data sharing. SAGE has the opportunity to promote education by working closely with partners and agencies to address confidentiality concerns and demonstrate how these concerns are managed. Others felt their organization could improve on marketing, especially through newer formats such as social media.

Two areas of improvement specific to SAGE based on internal and external interviews, and the document review were related to communication and linking data sets. In terms of communication, the ACs were not meeting as frequently as outlined in the documents and AC members felt they would benefit from more regular updates about SAGE. Further, external stakeholders stated that the ability to link across administrative, research, and community data would make SAGE unique in the field; however, communicating and making arrangements with data partners was slowing the process, particularly as it relates to administrative data. According to SAGE, there are currently barriers to releasing administrative data, but partnerships are being made with the government to move forward with the goal of eventually being able release the data and having administrative data available through SAGE.

Considerations and Potential Next Steps

A number of considerations and potential next steps for SAGE were provided by internal and external interview participants or outlined in the documents. A full list is available in Appendix D.

A number of suggested considerations by external interview participants were related to features they would find useful if they were data depositors or accessors. For instance, posting privacy policies online so people know how their data is protected, which would also encourage the use of SAGE. Posting online tutorials or support videos explaining how to extract data was also suggested.

Other considerations were associated with expanding collaborations and partnerships. These included other data repositories with different data, or private industry such as pharmaceuticals that also have a stake in health research and policy. Partnerships with private industry also offered the potential for alternative funding. SAGE documents discuss potential opportunities to collaborate with other data repositories.

Further suggestions from external interview participants included working more closely with community agencies to generate standards for metadata development, which could be facilitated in partnership

with the academic community. According to interviews with SAGE, discussions were in the works with community organizations to potentially offer services to make their data more useable and address their desire to understand their programs and services better based on that data. Though these suggestions offer several opportunities for expansion, one interview participants stated that SAGE should not try to do everything alone or put too much pressure on the initiative while it is getting off the ground.

According to interviews with SAGE, there is a need to stay relevant in a dynamic system. That includes maintaining momentum, building a community of support, keeping people engaged, and learning from others. One consideration moving forward is to incorporate qualitative data into the repository, though it has challenges of its own including small groups and identifiable data. Additionally, SAGE documents describe goals of aligning PolicyWise' and SAGE's strategic plan with the Government of Alberta's, which according to the CDCA Implementation Proposal (ACCFCR, 2013), "calls for 'Investing in Families and Communities' and 'Supporting Vulnerable Albertans'" (pg. 4).

<i>Key Considerations</i>	<i>Related Findings</i>
Include privacy policies on the SAGE/PolicyWise website	Interview participants said having privacy policies and instructions for extracting data easily available was useful
Upload tutorials or videos showing data accessors how to extract data	
Weigh the benefits and drawbacks of partnering with the private sector	Partnerships with the private sector can offer alternative funding opportunities and stakeholder views
Consider reviewing goals and linking to short-, medium-, and long-term	Data repositories have opportunities for growth and expansion in several areas; however the primary goal is to provide access to data

Conclusion

This is the first report of this evaluation and captures the initial work done. This document:

- Describes the context within which SAGE is implemented;
- Outlines current delivery and comparable leading practices; and
- Identifies and explains key contextual factors or dynamics that influence SAGE.

Based on interviews, and a document and literature review there was six overarching themes. The first was privacy and ethics, which addressed the main concerns and how they were being addressed. There were several established approaches to managing privacy and ethics, including SAGE's comprehensive policies and protocols.

Broad culture of data sharing and repositories indicated that cultural change is occurring, and though there are still barriers, a significant amount is being done. Ongoing education and awareness building were frequently cited as facilitators to supporting change and addressing barriers.

Goals, outcomes and impacts of data sharing captured how SAGE and others are working to increase access and availability of data, mainly through building relationships and strengthening and mobilizing capacity. This was particularly relevant for students and new researchers with fewer resources. Prioritizing data sharing, building relationships and strengthening and mobilizing capacity were facilitators of increased impactful work. Data sharing can increase meaningful knowledge and understanding, which has impacts for policy.

Operations significantly varied across interview participants; however, the processes put in place by SAGE generally align with the majority of practices, particularly of other established repositories. Likewise, SAGE had similar governance structures to other data repositories, with the exception of an additional AC, which was described as leading practice. Both ACs had accomplished a significant amount; however, there was some blurring between their respective roles. Additionally, both SAGE team members and external stakeholders said the ACs did not meet as frequently as they would like. Consideration for increased meeting frequency and communication would be beneficial.

Overall, interviews, the document and literature review indicate that SAGE has been working well. Several interview participants stated that SAGE was a relevant player in the field of data repositories, and the eventual ability to link research, community, and administrative data, along with a focus on children and families would make SAGE unique. Interview participants noted that the data already held with SAGE is clean and relevant. Additionally, SAGE was viewed as experts in the area of data privacy.

References

- The Alberta Centre for Child, Family and Community Research (ACCFRC). (2013). *The Child Data Centre of Alberta Implementation Proposal*. Retrieved from Internal Document Management.
- The Alberta Centre for Child, Family and Community Research (ACCFRC). (2016). *The Child Data Centre of Alberta Implementation Proposal*. Retrieved from Internal Document Management.
- Adair, C. (2012). *The Child Data Centre of Alberta Development Project*. Retrieved from Internal Document.
- Corti, L., Van den Eynden, V., Bishop, L., & Woolard, M. (2014). *Managing and Sharing Research Data: a Guide to Good Practice*. London: UK Data Service.
- Dove, E. S., Townend, D., Meslin, E. M., Bobrow, M., Littler, K., Nicol, D., et al. (2016). Ethics review for international data-intensive research. *Science*, 251(6280), 1399-1400.
- Economic and Social Research Council. (2013). *ESRC Framework for Research Ethics*. Retrieved from <http://www.esrc.ac.uk/files/funding/guidance-for-applicants/esrc-framework-for-research-ethics-2010/>
- Humphrey, C. (2006). *e-Science and the Life Cycle of Research*, University of Alberta. Retrieved from <http://datalib.library.ualberta.ca/~humphrey/lifecycle-science060308.doc>
- Kanous, A., Brock, E. (2015). *Contractual Limitations on Data Sharing*. Retrieved from <http://datacommunity.icpsr.umich.edu/sites/default/files/ContractualLimitationsonDataSharing150411-1.pdf>
- Kemmis, S., McTaggart, R. & Retallic, J. (Eds.). (2004). *The action research planner*. Karachi: Aga Khan University, Institute for Educational Development.
- Lee, J. E., Sung, J. H., Barnett, M. E., & Norris, K. (2015). User-Friendly Data-Sharing Practices for Fostering Collaboration within a Research Network: Roles of a Vanguard Center for a Community-Based Study. *Int J Environ Res Public Health* (2015), 13:1.
- Medical Research Council. (2016). *Data Sharing Policy*. Retrieved from <https://www.mrc.ac.uk/documents/pdf/mrc-data-sharing-policy/>
- Meslin, E. M., Rager, J. B., Schwartz, P. H., Quaid, K. A., Gaffney, M. M., Duke, J., et al. (2015). Benchmarks for ethically credible partnerships between industry and academic health centers: beyond disclosure of financial conflicts of interest. *Clin Transl Med*(2015),4:36.
- National Institutes of Health. (2003). *Final NIH Statement on Sharing Research Data*. Retrieved from <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>

- Organisation for Economic Co-operation and Development. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Retrieved from <http://www.oecd.org/science/scitech/38500813.pdf>
- Quinn Patton, M. (2006). *Evaluation for the Way We Work*. Retrieved from <https://nonprofitquarterly.org/management/646-evaluation-for-the-way-we-work.html>
- Research Data Canada. (2011). *Mapping the data landscape: Report of the 2011 Canadian Research Data Summit*. Ottawa: Author.
- Scottish Health Informatics Programme. (2010). *SHIP Guiding Principles and Best Practices*. Retrieved from http://www.scot-ship.ac.uk/sites/default/files/Reports/Guiding_Principles_and_Best_Practices_221010.pdf
- ZGM. (2016). *The Child Data Centre of Alberta Marketing and Communications Plan*. Retrieved from Internal Document Management

Appendix A

Introduction and Consent

Thank you for agreeing to participate in the “SAGE Evaluation Project”. We greatly value your time and feedback. The purpose of this evaluation is to support ongoing monitoring, adaptation, and improvement, as well as to evaluate the outcomes and impact of SAGE, specifically from October 2016 to now. This interview is to help inform our understanding of how SAGE was developed, how it operates, and what its successes and challenges are.

This interview will take approximately one hour. Participation in this interview is voluntary and you can end the conversation at any time or choose not to answer certain questions. Your answers are confidential and will only be used for project purposes.

We would also like to record the interview. The recording of our conversation will be kept on a secured, locked and protected site, and nobody outside the project will have access to it. Are you comfortable with this interview being recorded?

Do you have any questions for us before we get started?

Questions

1. How long have you been involved with SAGE/CDCA?
 - Can you please describe your role and responsibilities with SAGE?
2. Can you describe how SAGE works? (In other words, what is your elevator speech?)
 - Is there a particular program model that SAGE is modelled after?
 - What do you see as the main outcomes and impact of SAGE?
 - Does SAGE provide additional resources/services alongside data sharing? If so, what are they?
3. Who is the target population for using SAGE (both for data deposit and access)?
 - How is the use of SAGE promoted (both for data deposit and access)?
4. In terms of **depositing** data:
 - How do people/organizations deposit their data?
 - Are there specific criteria around who can deposit data? How do you ensure that the data is ‘good quality’?
 - Are the original contributors involved when their data is used? If so, what does this involve?
5. In terms of **accessing** data:
 - How do people/organizations access the data?

- Are there specific criteria around who can access data?
6. How has SAGE evolved over time?
 - Why do you think those shifts occurred?
 - Were there any contextual factors that influenced the shifts in SAGE?
 - What lessons have you learned along the way?
 - When the name was changed, were there any other corresponding changes?
 7. Could you please tell us about the strategic and operational advisory committees?
 - When did they start?
 - How was membership determined?
 - What purpose do they each serve?
 - How do they operate? (Ex. frequency of meetings, decision making process)
 - What has worked well and/or not well with these ACs?
 8. In planning documents created for SAGE/CDCA, a few potential issues were identified. Have these been addressed? And if so, how?
 - Crediting contributors for their data
 - Concerns about privacy
 - Investigators' loss to exclusivity of the data
 9. What (if any) other initiatives do you see as similar to SAGE?
 - How do you see SAGE in comparison to other similar initiatives?
 - What do you think are the key ingredients to SAGE's success?
 10. Are there people who you would consider to be thought leaders in this area? If so, who?
 - Why would you considered her/him to be a thought leader?
 11. Is there anyone else you think we should talk to for this evaluation project?
 12. Reflecting on SAGE overall, can you tell us about what you think has been working well and what has not been working well?
 - Are there any barriers to SAGE's success? If so, what are they?
 - What (if anything) would you say could be improved?
 13. Is there anything else that you feel is relevant to the evaluation of SAGE that we have not discussed?

Conclusion

Thank you for your time today. Hearing from you has been an important part of this project. Our next steps include continuing to conduct interviews with other stakeholders and to review SAGE-related

documents. After this, the project will continue with reviewing leading practices and comparative models on data repositories, including exploring academic and grey literature and conducting interviews with like-minded program staff.

Appendix B

Introduction and Consent

Thank you for agreeing to participate in the “SAGE Evaluation Project”. We greatly value your time and feedback. I/we work for PolicyWise for Children & Families, which is a provincial not-for-profit organization that exists to improve well-being by leading, creating, enabling, and mobilizing research and evaluation for evidence-informed policy and practice. Within PolicyWise, we have a data repository called **SAGE- Secondary Analysis to Generate Evidence**. We are conducting an internal evaluation of this initiative.

The goal of our evaluation is to support ongoing monitoring, adaptation, and improvement, as well as to evaluate the outcomes and impact of SAGE. The purpose of this interview is to help inform our understanding of the different perspectives of users either accessing or depositing data. .

This interview will take approximately one hour. Participation in this interview is voluntary and you can end the conversation at any time or choose not to answer certain questions. Your answers are confidential and will only be used for project purposes. With your permission, we would like to record the interview. The recording of our conversation will be kept on a secured, locked and protected site, and nobody outside the project will have access to it. Are you comfortable with this interview being recorded?

Context

1. What is your role?
 - How long have you been involved in this role?
2. Do you have experience with data sharing (depositing and/or accessing)?
 - If yes, what has been your experience?

SAGE

3. Are you familiar with SAGE?
 - How did you become familiar with SAGE?
 - Can you describe your understanding of SAGE?
4. Have you had previous involvement with SAGE/PolicyWise in some form? If so, what did that involve?
5. Have you ever used SAGE? (If yes, continue with probes)
 - What was your experience like using SAGE?
 - What worked well?
 - Did you come across any challenges?
 - If so, what could be done to improve the process?
 - Is there anything that you would you change about SAGE? If so, what?

- What (if any) support did you receive?
 - What support would be most beneficial to receive from the SAGE team?
 - Once you understood the function of SAGE, did you find it valuable?
6. What do you think are the key ingredients to SAGE's success?
7. What (if any) other initiatives do you see as similar to SAGE?
- How is SAGE different from other initiatives?
 - What do you think is SAGE's niche?

Data repositories in general

8. Reflecting on data repositories overall, how would you describe the current state of data repositories? (ex. culture, attitudes, trends, etc.)
9. Can you tell us about what you think has been working well and what has not been working well?
- What (if anything) would you say could be improved?
10. Are there any barriers to success? If so, what are they?
- How could they be addressed?
11. Do you see any current gaps in data repositories?
12. Are there any key resources you would recommend for us to consider for this project?
13. What do you think would encourage data sharing? (for yourself, for others with the same kind of data, for those with other kinds of data)

AC-related questions (as needed)

14. Please describe your experience on the Operational Advisory Committee.
- What support/guidance is provided by the AC for SAGE?
 - In your opinion, what is working well? What is not working well?
 - How do you feel about the AC's meeting frequency?
 - Do you think that the way the current AC operates fits appropriately with what SAGE needs?

Conclusion

15. Is there anything else you feel is relevant to consider for data repositories that we have not discussed?

Thank you for your time today. Learning from you has been an important part of this project!

Appendix C

Introduction and Consent

Thank you for agreeing to participate in the “SAGE Evaluation Project”. We greatly value your time and feedback.

I/we work for PolicyWise for Children & Families, which is a provincial not-for-profit organization that exists to improve well-being by leading, creating, enabling, and mobilizing research and evaluation for evidence-informed policy and practice. Within PolicyWise, we have a data repository called **SAGE – Secondary Analysis to Generate Evidence**. We are conducting an internal evaluation of this initiative.

The goal of our evaluation is to support ongoing monitoring, adaptation, and improvement, as well as to evaluate the outcomes and impact of SAGE. The purpose of this interview is to help inform our understanding of how other data repositories operate, hear about your successes and challenges in this area, and learn from your experience.

This interview will take approximately one hour. Participation in this interview is voluntary and you can end the conversation at any time or choose not to answer certain questions. Your answers are confidential and will only be used for project purposes.

With your permission, we would like to record the interview. The recording of our conversation will be kept on a secured, locked and protected site, and nobody outside the project will have access to it. Are you comfortable with this interview being recorded?

Context/their initiative

14. What is your role in your initiative?
 - How long have you been involved in this field?
15. Can you tell us a bit about your initiative?
 - a. How does it operate? (depositing, accessing, etc.)
 - i. How much flexibility do you allow for in these processes?
 - b. Has your initiative had to address the following issues? And if so, how?
 - Concerns about privacy
 - Investigators’ loss to exclusivity of the data
 - Crediting contributors for their data
 - What lessons have you learned about data repositories and/or secondary data use in general?
 - How does your initiative stay up to date/current with leading practices?
 - What is your governance like? (ex. do you have an advisory committee, board, etc.)
16. What do you see as the main outcomes and impact of your initiative?

- a. What do you think are the key ingredients to its success?
17. Is there a particular program model that your initiative is modelled after?

Data repositories in general

18. Reflecting overall, how would you describe the current state of data repositories? (ex. culture, attitudes, trends, etc.)
19. Can you tell us about what you think has been working well and what has not been working well?
- What (if anything) would you say could be improved?
20. Do you see any barriers to success? If so, what are they?
- a. How could they be addressed?
21. Do you see any current gaps in data repositories?
22. What do you think would help to encourage data sharing?
- a. What strategies does your initiative use to encourage secondary data use?
23. Are there any key resources you would recommend for us to consider for this project?

SAGE

24. Are you familiar with SAGE? If so:
- How did you become familiar with SAGE?
 - Can you describe your understanding of SAGE?
 - Have you had previous involvement with SAGE in some form? If so, what has that involved?
25. What (if any) other initiatives do you see as similar to SAGE?
- How do you see SAGE in comparison to other similar initiatives?
 - What do you think are the key ingredients to SAGE's success?

AC-related questions

26. Please describe your experience on the Operational Advisory Committee.
- What support/guidance is provided by the AC for SAGE?
 - In your opinion, what is working well? What is not working well?
 - How do you feel about the AC's meeting frequency?

Do you think that the way the current AC operates fits appropriately with what SAGE needs?

Conclusion

27. Is there anything else that you feel is relevant to consider for data repositories that we have not discussed?

Thank you for your time today. Learning from you has been an important part of this project!

Appendix D

	Key Considerations	Related Findings
Administrative Data	Work with ministries to develop policies and processes to protect data	
	Demonstrate research participant willingness to share data	Privacy and individual identification is the greatest concern for administrative data
	De-identify data using numerical codes or algorithms even when not required by the data partner	
	Facilitate conversations with other data repositories and governing bodies about potential ethical issues and discuss next steps	Ethical questions have been raised about actionable research findings and the appropriate response
Research Data	Work with researchers early in the research process to ensure proper ethics and consent are obtained for data sharing	Ethics and consent are concerns for researchers sharing data. Consent for data sharing is often not obtained at the beginning of the research process
	Assist with developing ethics and consent submissions that leave open the possibility for data sharing	Research ethics boards conservatively interpret ethics and consent forms
Community Data	De-identify community data using numerical codes or algorithms	Community agencies are concerned about protecting client privacy
	Offer support to community agencies by providing training on data collection, analysis, and data sharing	Agency staff may not have the expertise to collect or analyze data
Data Sharing & Legislation	Develop a streamlined data sharing agreement that can be applied to every project while leaving room for adaptation	Data sharing agreements for each project is time consuming and slows the process down
	Assist researchers with drafting clearly defined data sharing agreements	Unintentional data restrictions may result from unclear data agreements produced by researchers
Managing Risk	Implement data accessing protocols and vet potential data accessors	Risk can be effectively managed using a variety of techniques
Culture Change	Work with PolicyWise and SAGE grant recipients to encourage data sharing	Funders are recognizing the benefit of data sharing and may require funded researchers to share their data (i.e., NIH, ERC)
	Continue promoting data sharing through grants, training and conferences while considering areas for expansion	Education and promotion targeted at data users, funding bodies, community organizations, and institutions supports cultural change related to data sharing
	Establish champions of SAGE that will promote the initiative, and educate other through mainstream channels and social media	

Supporting Culture Change and Addressing Barriers	Demonstrate the usability of data by advertising projects completed using existing data	
	Promote methods for secondary analysis among students and university faculty	There is a perception that primary data is superior, or researchers are unaware secondary data is available to them
	Support data storage for researchers once they leave the field to ensure long-term preservation and access	
	Encourage funders to provide more funding opportunities for researchers collecting data for deposit or conducting secondary analysis	The cost of producing data with large enough samples is high, and though secondary data analysis is cheaper, it is still costly and time consuming
	Collaborate with other data repositories and governments to promote and facilitate the development of frameworks or policies that support data collection, preservation, and sharing	There is a lack of international ethics frameworks, which hinders international data collection and data sharing.
	Develop standard ethics and consent language around data sharing	Additionally, Canada does not have a data management plan for data collected using public funds
Prioritizing Data Sharing Processes	Involve research participants in larger discussions about consent and data sharing	
	Support researchers early in the research process starting with funding proposals, if possible, to leave open the opportunity for data sharing	Clear research plans that include potential data sharing mitigate ethics and consent concerns and account for changes in roles or responsibilities throughout the research/data lifecycle
	SAGE is engaged in several leading practices that others would benefit learning from. Continue participating in opportunities to share SAGE's approach and learn from others	Overall goals of data repositories included developing better storage and preservation approaches, and contributing to leading practices on data sharing
Building Relationships	Publish articles, reports, and other documents	
	Continue engaging and building relationships with universities, community agencies, and potential data accessors or depositors	
	Facilitate collaborations or partnerships between students and junior faculty and researchers interested in depositing their data for secondary analysis	SAGE has developed many partnerships and collaborations, particularly in the area of government ministries and ethics bodies

	Work with cross-sectoral partners to develop privacy standards, MOU's, etc. to facilitate cross-sectoral data sharing Represent the needs of other sectors through existing relationships with government ministries Continue leveraging relationships developed by PolicyWise	Cross-sectoral data sharing and partnerships are beneficial and lead to a more holistic understanding
Strengthening & Mobilizing Capacity	Consider offering preservation packages as an added service for researchers without the tools necessary	The overall goal of data sharing is to increase access and use of existing data
	Using SAGE's expertise to offer courses and training on data management and curation to students, junior faculty, or other researchers	There are few skilled workers in the area of data management and curation in Canada
Increasing Impactful Work	Continue to work with government ministries to merge or incorporate data into SAGE	Linking data offers an accurate picture of the services needed or being accessed by people, which can be used to improve policy decisions
Promotion and Visibility	SAGE is continuing to explore the best approach to promotion and visibility. A combination of broad approaches (i.e., conferences, social media, publications) and targeted approaches (i.e., grants, reaching out to previous researchers) is beneficial	Adapt promotion based on the changing needs of users and technology
Data Requirements	Accepting de-identified data may ease privacy concerns associated with administrative data	The majority of data repositories holding administrative data only accepted de-identified data or restricted access within their organization even when it was not required by the data repository
	SAGE currently has a focus on social sector data related to children and families; the scope of SAGE can be narrowed, maintained, or expanded based on the needs of data depositors and accessors	Data repositories usually offer either general data collections or specific data collects, each of which have pros and cons
Data use Requirements	Data repositories holding administrative data also had policies related to internal staff access; since holding administrative data is a goal for SAGE, consideration for who within SAGE/PolicyWise will have access may be useful	Most data repositories, including SAGE, have similar data access processes, including requests or applications, reviews, and data sharing agreements; there was variation in how data is shared
Other Operations	As SAGE grows the level and type of support or flexibility offered may need to be adapted	The majority of repositories offered some support, recognizing depositors and accessors have different levels of experience and support needs; however, experimentation with the level of support and resources put towards it was common

Governance	Some repositories had or were considering cost recovery mechanisms to supplement funding	Maintaining repositories has associated costs
	A number of data repositories had agreements with other organizations or repositories to hold the data in the event the repository ceased to exist	Building trust with data repository users includes reassurance that it will be around over time, but having a contingency plan in case
	A larger council or committee to represent interests of Canadian data repositories was suggested	There is limited representation of data repositories on a national or international level, contributing to inconsistent policies and legislation
	Implementing two committees places SAGE ahead of other data repositories	Developing a governance committee or council was good practice
	Increasing the frequency of meetings, sending out regular updates, and meeting in-person occasionally may improve communication and the overall experiences of AC members	The OAC and SAC were not meeting as frequently as anticipated
Considerations and Potential Next Steps	Include privacy policies on SAGE/PolicyWise website	Interview participants said having privacy policies and instructions for extracting data easily available was useful
	Upload tutorials or videos showing data accessors how to extract data	
	Weigh the benefits and drawbacks of partnering with the private sector	Partnerships with the private sector can offer alternative funding opportunities and stakeholder views
	Consider reviewing goals and linking to short-, medium-, and long-term	Data repositories have opportunities for growth and expansion in several areas; however, the primary goal is to provide access to data