SUMMARY REPORT 2018

Using Data About You for Research: Who, How, and Why

A public deliberation with British Columbians

Table of contents

Executive Summary

A public deliberation event, ***Using Data About You for Research: Who, How, and Why***, was held in Vancouver, British Columbia on April 7-8 and 21-22, 2018. It involved British Columbians with a range of age, sex, area of residence, income, and ethnicity characteristics, all of which was intended to maximize diversity in life experience and opinions about the deliberation topic. This report summarizes the purpose, the methods, and the policy recommendations that were developed by the participants.

The purpose of this event was to develop informed and civic-minded public advice from British Columbians regarding the use of linked data for research. Research using linked data is increasingly conducted by both public (e.g., academics, government Ministries) and private organizations (e.g., pharmaceutical and device industries, technology companies). While using linked data for research has the potential for discoveries that positively impact society, it also raises concerns relating to illegitimate use, privacy, and security (e.g., identity theft, marginalization). Balancing the potential benefits from research and the associated risks will require making trade-offs that should be informed by the public's perspectives and interests.

Over the course of four days and in preparation for the event, participants read written material, heard expert and stakeholder speakers, explored their own perspectives, and considered how to balance the diverse interests related to research using linked data. Participants formulated and voted on nineteen recommendations. Voting was a tool to help identify, explore and document the reasons behind participants' agreeing, disagreeing or abstaining with each recommendation.

Key findings from the deliberation

Participants were supportive of research using linked data because of the value it provides to society. They expressed a desire to see the data access request process made more efficient in order to facilitate more research, as long as there are adequate protections in place around security and privacy of the data. Notably, participants expressed these positions at an event that happened during a period of heavy media coverage of the Cambridge Analytica/Facebook allegations of inappropriate data usage; given the circumstances, heightened concerns and a more conservative position would not have been surprising.

Participants were encouraged to start from a blank slate and make the recommendations they thought represent good practice without respect to what is or is not in place currently. They developed recommendations that address the following topics. The full report provides detail on the degree of support and reasons behind voting for each recommendation

Recommendations regarding the governance of linked data

- Financial investment in linked data sets
- Efficiency and fast tracking in the approval process
- Limited involvement of commercial entities
- Need for an oversight committee for all aspects of data sharing

Recommendations regarding the security and review process for linked data and the results from their research

- Review of the results and publication of research using linked data
- Development of best practices and guidelines for storage and access to linked data
- Need for an ethics review process for research proposals
- Need for an independent body to oversee research data requests
- Need for an independent body to monitor the use of data by researchers

Recommendations regarding the responsibilities of researchers and Data Stewards

Data Stewards:

- Development of training and certification for Data Stewards
- Development of standard policies and procedures for managing data access requests
- Data Stewards have limited need to monitor research results

Researchers:

- Requirement for contracts outlining confidentiality requirements
- Research responsibility to vulnerable populations
- Development of a certification program on how to use linked data securely

Recommendations regarding public involvement

- Public disclosure of the requests for access to linked data
- Role of transparency and disclosure in public consultation

Introduction

Why deliberate on linked data for research?

The public deliberation event, *Using Data About You for Research: Who, How, and Why*, addressed issues around the use of linked data for research. The event was held in Vancouver, British Columbia, over four days including April 7-8 and 21-22, 2018. The project was organized by Population Data BC and a research team from the University of British Columbia and the University of Guelph. A grant from the Canadian Institutes of Health Research supported the project (Principal Investigator, Dr. Kimberlyn McGrail).

Research using linked data is increasingly being conducted by both public (e.g., academics, government Ministries) and private organizations (e.g., pharmaceutical and device industries, tech companies).[1] This trend is made possible by the increasing availability of data in digital form, technical advances in linking complex data sets, and scientific advances in analyzing the resulting linked data. Research using large and often complex linked data sets can lead to new insights that inform policy or service delivery that affects the health of individuals, the efficiency of processes and finances, and the distribution of resources.

What is a public deliberation?

A public deliberation is a community discussion that involves citizens, policy makers, and experts formulating recommendations that considers all perspectives.

Researchers desire more access to linked data sets, including data from new sources and types of data collections, e.g. patient-reported information, genomic information, and data from wearable devices, social media and so on. Current policies and practices around the sharing of linked data will need to adapt to this new reality, for example understanding who can have access to what under what circumstance.[2]

While researchers have clear ambitions, the use of linked data raises other concerns, such as privacy and security as well as improper and unethical use.[3] These concerns are very present in the public mind, as controversies involving Facebook, Cambridge Analytica, and privacy breaches are covered regularly in the media.[4] Notably, this event was held during high media coverage of the unfolding of Facebook/Cambridge Analytica activities.

What does it mean to link data?

Linked data are collections of data that combine two or more data sets, such as data sets from ICBC and BC Vital Statistics. Many research projects request access to far more than two data sets that span multiple organizations.

The main tension with the use of linked data for research is that it holds both great potential to benefit society and raises significant concerns around privacy (e.g., identity theft), security (e.g., illegitimate use of data), and ethical behaviour (e.g., risks of stigmatization through data manipulation). Working with this tension will necessarily require making trade-offs. These trade-offs must be informed by expert knowledge but should also be informed by the public's perspectives and interests. Developing these and developing policy recommendations that reflect them was the purpose of this deliberation.

Principles of public deliberation

Public deliberation events are informed by political theory on deliberative democracy.[5] They are based on the idea that it is important to have citizen input on issues that are controversial or a source of concern. The purpose of public deliberations is not to convince participants of any given position or bring them to consensus on the issues being discussed, but for participants to deliberate among themselves and reach collective statements or recommendations. The goal is to encourage participants to work together to develop recommendations for policy that accommodate their varied perspectives. As such, public deliberations offer recommendations that, if followed, can enhance the democratic legitimacy of programs, actions, and decisions.[6]

The main premise of public deliberations is that, in spite of differences in opinion and interests, members of society need to find common rules and practices consistent with this diversity. Deliberation events are distinct in that they invite the public into active discussion about important societal issues. They do so recognizing that individual members of society are experts in their own lives and have important things to say about policy.

Public deliberations can be distinguished from public consultations by the depth and length of the discussions, the amount of relevant information provided to and by participants, and how the participants themselves create the recommendations.[7] Public consultations often collect participant views, whereas deliberations are intended to create collective recommendations that reflect how participants think their diverse interests are best accommodated. The form of public deliberation we used occurs over multiple days and invests a significant amount of time to prepare and provide relevant materials to participants.[8]

Methods

Participant recruitment

Twenty-eight British Columbians initially participated in the deliberation. The number of participants in the final stage of the event decreased from 28 to 23, as a result of personal circumstances (e.g., health, emergency scheduling issues).

Participants were selected to reflect the diversity of residents in British Columbia in terms of age, sex, income, ethnicity and geographic area of residence. Diversity in demographic and regional characteristics was intended as a way to maximize diversity in life experience and perspective. The only group we explicitly excluded were people who work as privacy professionals; in public deliberation, it is best to avoid inclusion of people considered experts as other deliberation members tend to be deferential to that expertise, which undermines the deliberation process.[9]

To recruit participants, we worked with a marketing research company, CRC Research, which provided initial lists of potential participants, whom we contacted and invited to attend. The deliberation was held on the weekend to facilitate attendance. The costs of travel, meals and accommodations were covered by project funding. Each participant also received a $150 honorarium per day of attendance.

We were conscious of recruiting participants from all age groups and made special attempts to include participants aged 18-24 years as previous deliberations indicated this group is difficult both to recruit and to maintain. We also make specific effort to recruit individuals who identified as Indigenous, as there are distinct norms and practices around data and data sharing in Indigenous communities that were important to reflect in the deliberations. Though we did have good representation of people 18-24 years, we were able to recruit but not maintain representation from Indigenous participants.

Details on the demographic composition of the participants are in Appendix A.

Informing participants

Participants were not required to have prior knowledge about the use of linked data for research. They were encouraged to bring their opinions, values, and ideas about data to the deliberation. To support participants' discussions and to ensure that they all had the same base of knowledge, we prepared materials specifically for the deliberation. These consisted of an information booklet[10] that was provided two to three weeks in advance of the event, and presentations from expert speakers on the first day of the event. The goal of these materials was not to make pubic members into lay experts, but to provide a broad range of views on the issues central to the deliberation. This supports individuals to participate confidently in the discussions, to feel comfortable expressing their views, and to be able to engage with and respond to other participants' contributions.

The deliberation booklet was developed by the research team using expert knowledge and academic literature. The booklet described what linked data are, how they are collected, what regulations need to be followed to share them, and current issues and concerns surrounding their use. A glossary provided definitions for technical terms. Participants and expert speakers were provided with a digital and physical copy of the booklet.

At the event, the participants heard presentations from expert speakers representing a range of perspectives, from data stewards, privacy advocates, researchers, and community stakeholders. The speakers included Laura MacDougall (Director of Public Health Analytics at the BC Centre for Disease Control), Sherri Pooyak (Community-Based Research Manager with the Canadian Aboriginal AIDS Network), Teri Thorson (Spinal Cord Injury BC), Michael Vonn (Policy Director at the BC Civil Liberties Association), and Ryan Woods (Scientific Director of and data steward for the BC Cancer Registry). The speakers all took part in an informal panel and answered participant questions.

Deliberative process

The event format followed deliberative public engagement methods developed by Burgess, O'Doherty et al.[11] which have been used in previous deliberation events.[12–14] The event occurred over two non-consecutive weekends (i.e., a total of four days) in April 2018 and was run by facilitators trained specifically for this type of public deliberation. The time between the weekends was to give the participants the opportunity to return home and reflect on discussions. Participants were encouraged to use their time away to discuss the topic with their family and friends.

Participants met in both small group and large group settings during the event. There were four small groups that consisted of 6-8 participants each. The intent of the small group deliberations was to encourage participation by all attendees, as well as generate a broad range of viewpoints on the topic of discussion. The intent of the large group deliberations was to work towards bringing the diverse views articulated in the small groups to bear on the issues that the large group considered important, and then to craft policy recommendations on the use and sharing of linked data for research, as well as to capture the reasons for and against the recommendations.

The first day was focussed on providing participants with information about using and sharing linked data for research, as well as introducing them to the process of public deliberation. Expert speakers gave their presentations and answered the participants' questions. Participants were also told they could ask for additional information and the research team would seek this information and present it to the group on subsequent days.

The second and third days were used to discuss the deliberation questions. For each question, the participants first discussed the issues in their small groups. They then returned to the large group to discuss issues and different perspectives raised in the small groups. The facilitator helped formulate the points of discussion into preliminary statements. Participants then worked together to edit the statement until it represented a collective position that could be used as a policy recommendation. With that complete, the participants voted on the statement, indicating if they were for or against the statement, or if they were abstaining from the vote. Participants were asked to provide the reasoning for their position. This diversity of perspectives was explored and documented.

The fourth day was used to summarize the group's recommendations. They also engaged in a ratification process to ensure that there was still support for all the recommendations. This was an opportunity for participants to change their vote in light of new information and changing perspectives. The deliberation concluded with a panel of experts who work with data and are in a position to influence policy around sharing linked data for research. The panel members included

Hayden Lansdell (Executive Lead of the Integrated Data Division within the BC Government), Caitlin Lemiski (Senior Policy Analyst with the BC Office of the Information and Privacy Commissioner), Matt Reed (Acting Executive Director of the Privacy, Compliance and Training Branch for the BC Government), and Suzanne Vercauteren (Director BC Children's Hospital BioBank). The final session allowed the experts to hear the recommendations developed by the participants and for a discussion between the experts and the participants.

All proceedings were audio recorded and transcribed. Detailed analysis of the transcribed proceedings is underway. As this is a research project, all procedures were reviewed and approved by the UBC Behavioural Research Ethics Board.

See Appendix B for the event schedule.

Deliberation questions

The deliberation was structured around three questions and an exercise using plausible data access request scenarios. The deliberation questions and the scenarios were developed by the research team.

Participants were asked to discuss the deliberation questions within small and large groups. They were encouraged to move towards collective positions that could be represented as civic-minded solutions in the large groups, taking into consideration the different perspectives raised in the small group.

Participants were also asked to provide feedback on the scenarios and to indicate any concerns about them and what changes they would recommend. This process greatly aided the participants' ability to grasp the trade-offs in sharing linked data.

The deliberation questions and scenario exercise used in the deliberation event in chronological order

1. What is important information to consider when approving access to and use of linked data?

2. When is it justified to grant access to linked data, and what measures are important to reduce risks?

3. Working with scenarios: applying the discussions from the previous weekend to work out trade-offs and recommendations

4. What processes would make the assessments of risks and benefits from the use of linked data trustworthy?

Participants' recommendations: preliminary results

The participants' recommendations are grouped into four categories (they are not in the order they were developed in the event):

1. Recommendations regarding the governance of linked data;
2. Recommendations regarding the security and review process for linked data and the results from their analysis;
3. Recommendations regarding the responsibilities of researchers and Data Stewards;
4. Recommendations regarding public involvement.

Although the participants voted on the recommendations, the numeric votes should not be over-interpreted. For example, while some of the "against" votes were opposed to the recommendation, other "against" votes had issues only with the wording of the recommendation, and sometimes even desired the recommendation to take a stronger position. The voting was primarily used as a tool by the facilitator to assess agreement or disagreement. After each vote, the facilitator could identify individuals who disagreed or abstained and ask them to explain their positions. In cases where there were disagreements on the recommendations, we made sure to record a clear articulation of the disagreement as well as the reasoning behind it, since disagreements and reasons for them can reveal differences in acceptable trade-offs among the public.

In each section, the recommendations are ordered by the degree to which the participants converged in their positions.

Representation of the participant votes

Participant votes are represented in color-coded pie charts next to each recommendation.

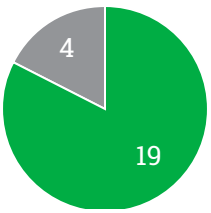
Green indicates "for" votes, orange indicates "against" votes, and grey indicates "abstain" votes.

The number of votes for each category is shown on the pie chart.

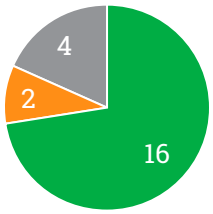
Recommendations regarding the governance of linked data

1. Develop a plan to make the data linkage approval process more efficient, without compromising the evaluation process.

This recommendation resulted from the participants discussing reports that it takes a long time for Data Stewards and researchers to secure approvals for access linked data. Participants were concerned about delays unnecessarily creating barriers to research being conducted. Those who **abstained** were concerned that the increased speed would result in corners being cut resulting in reviews that are superficial.

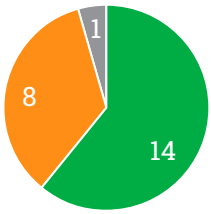


2. It is important to invest in a collection of linked datasets to promote efficient research while enhancing privacy protection.



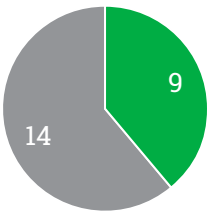
Participants recognized current efforts to protect privacy as well as the potential benefits of research from linked datasets. They expressed that both could be enhanced by increased investment of funds. Those who **abstained** wanted more information about who would have access to the datasets, whether there would be consistency in rules for access and data protection across organizations, and whether efficiency and privacy would be at cross-purposes. The participants who were **against** the recommendation were concerned that creating linked data sets implied centralizing information in one place, that this might imply having one single person in charge of making all related decisions, and that there would be uncertainty about what information would be contained in the data, all of which could increase risk.

3. Policy makers should establish categories that identify requests that require different paths or speed for review, e.g., fast-track for urgent research priorities.



This recommendation resulted from discussing the need for research to be conducted on urgent issues and how to fast-track data access requests in these circumstances. Participants were divided on the need for policy makers to establish categories to fast-track data access requests. In this case, the participants were not referring to the Data Stewards' daily activities. That is, most were not talking about whether the Data Stewards should have a faster review track for simpler applications. Instead, they were referring to the ability of policy makers to fast-track projects by declaring emergencies. The example that was brought up was the opioid crisis, and how it could benefit from having a policy maker to direct Data Stewards to prioritize the relevant applications. Those who **abstained** disliked the wording of the recommendation and wanted more specifics on logistical details like timelines. Those who were **against** felt that categories were unnecessary as policy makers would already act to fast-track data access requests for emergencies.

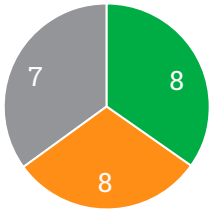
4. If a commercial entity funds research with linked data, it should not be involved in the production and review of that research.



Participants were concerned about commercial entities being involved in research and had reservations about how their profit motive would affect the use of linked data and the research results. However, participants did not converge on a position to articulate a recommendation. They instead developed a statement that captured some aspects of their position. Those who voted in **favour** indicated a desire for commercial entities not to have direct access to data. Those who **abstained** explained that if corporations contribute money, then they should have some benefit from the results of the research. They also were worried that if corporations were restricted from benefitting from the research, then they would have no reason to collaborate with universities. They also said that commercial entities using data was inevitable, so a recommendation banning them would not be useful. They indicated that there should be more clarity on what the involvement of commercial entities would entail.

5. There should be a committee or governing body with authority to:

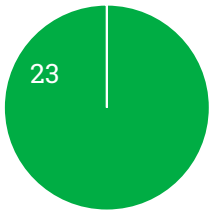
1. ***Provide oversight and investigation for breaches and/or harms***
2. ***Apply penalties or other consequences***
3. ***Develop policies to mitigate the potential for future breaches and/or harms***
4. ***Intervene when Data Stewards disagree***
5. ***Develop and operate an appeals process***
6. ***Provide certification for Data Stewards***



This recommendation resulted from participant concerns that there was a lack of overall guidelines and regulations governing access to linked data and the interactions between Data Stewards. They discussed cases where Data Stewards disagreed on data sharing policies and how there was no clear mechanism to resolve the disagreements. Participants suggested the creation of an independent committee and developed a list of responsibilities for it. Ultimately, they were unable to reach convergence on the recommendation. Those who **abstained** agreed in principle for the creation of committee but were concerned that the responsibilities as stated in the recommendation would be too wide-ranging and had the risk of being politicized. Some also expressed a general concern that there were too many committees already. Those who were **against** were also concerned that the scope of the committee was too large, and that there was a risk that its powers could be abused (e.g., to "quash research").

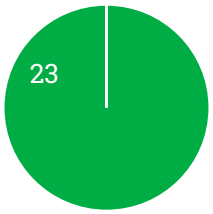
Recommendations regarding the security and review process for linked data and the results from their research

6. Scientific review of the research proposal should be performed by an independent party.



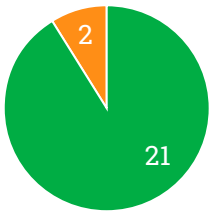
Participants were impressed with the potential of using linked data for research, but they were concerned about whether the research being conducted and the analysis of the results was being done correctly and by competent individuals. The recommendation for an independent scientific review was a result of discussing mechanisms to ensure that the assessment of the science behind the research proposal is done properly by qualified parties.

7. There should be best practices and guidelines for secure storage and access to linked data.



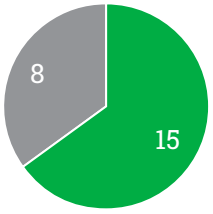
Participants repeatedly expressed concern about the security of linked data and wanted to know that access to the data was secure and would maintain privacy. The recommendation was a result of this discussion.

8. Results and publications of linked data research must be reviewed to ensure that they are justified by the analysis of the data.



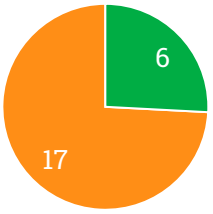
Participants were concerned that research results could have negative impacts or be misleading if not interpreted properly. They recommended that this concern could be addressed by ensuring that the results and publications are reviewed by an independent party. It is important to note that there were differences in how the participants interpreted this recommendation. Those who were **for** the recommendation believed the review of the results and publications would be completed by the peer review process. Those who were **against** believed the review is already conducted by Data Stewards and that the recommendation is redundant.

9. The proposed research and data access should be reviewed by an independent ethics committee to ensure benefits outweigh potential harms (e.g. potential for re-identification, stigma).



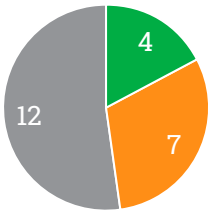
Participants were concerned about the risk that research may have harmful impacts on the populations they study, particularly in the case of vulnerable and marginalized populations (e.g., children, First Nations communities). They recognized that harms may result unintentionally, perhaps even without the researchers' knowledge. This discussion resulted in the recommendation for an independent ethics committee. Those who **abstained** were generally concerned about another procedure that would slow down the data access request; they also wondered if there was already an independent ethics review in place. They wanted the ethics review to apply to all linked data requests, not just those identified as "research."

10. Research results should be reviewed by a qualified independent party to reaffirm the original purpose of the research.



This recommendation was also part of the discussion that research may have potential harmful impacts. To address this, participants proposed that research results and the use of the linked data should be monitored to see if they followed the original intent of the research proposal. The proposal did not receive strong support. Those who were **against** the recommendation expressed that the committee would be redundant and that they were not clear on how the committee could enforce their reviews. There was also concern that the committee may lead to too much oversight of research. This related to concerns that the committee would not be useful, since the direction of research is not predictable, with the understanding that what researchers "do" is necessarily sometimes different from what they "propose".

11. An independent party should assess requests for data to be sure that the data are necessary to conduct the research.

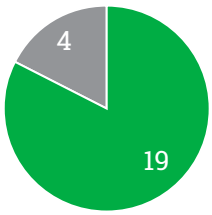


This was the final recommendation resulting from the discussion on the potential negative impacts of research using linked data. Here, participants were concerned that researchers may be receiving data that was unnecessary to complete their research; since researchers would have access to an increased amount of sensitive data, it was suggested that there was a higher risk of inappropriate usage or of a security incident. Participants did not converge on a position on whether it was necessary to ensure that the requested data matched the needs of the research proposal. Those who were **for** supported the general idea of monitoring the data are being used properly. Those who **abstained** said that while they supported the idea, they felt like it was a task that was already being done. Those who were **against** felt that this was redundant and the Data Stewards were already doing this. Participants explained that if the Data Stewards were qualified for their jobs, then there should be no need for an independent review.

Recommendations regarding the responsibilities of researchers and Data Stewards

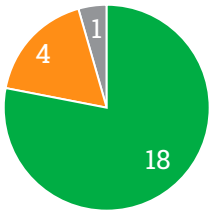
Data Stewards

12. Data stewards should have standard training or certification to ensure appropriate expertise for their role.



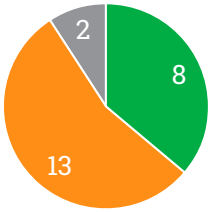
This recommendation was developed after discussing that Data Stewards do not have standardized training to do their role, resulting in variation in the way that data access requests are processed. Participants were concerned that the lack of standardization would decrease the efficiency of processing data access requests. Those who **abstained** were concerned with excess certifications that are not necessary. They were also concerned about whether it would be possible to standardize practices across different organizations.

13. Data Stewards should have standard policies and procedures to guide their work and there should be a certifying body to maintain them.



This recommendation was an extension of the discussion around the lack of standardized training for Data Stewards, which also covered a lack of standard policies and procedures. Participants were again concerned that non-standardized policies and procedures would result in inefficiencies. They raised the possibility that those policies and procedures could be developed and maintained by a certifying body. Those who **abstained** felt that the research institutions should be responsible for adhering to appropriate policies and procedures. Those who were **against** felt that the role of a certifying body was unclear and were also generally against certification boards.

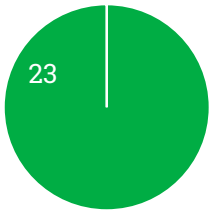
14. Research using linked data must be monitored by Data Stewards to ensure data are used in accordance with the original request.



This recommendation was related to previous recommendations (recommendations 8-11) and concerns around the potential impacts of research results and the misuse of linked data by researchers. Participants continued their discussion on how the use of linked data by researchers could be monitored and wondered if this was a responsibility that the Data Stewards should shoulder. Participants did not converge on a position for this recommendation. Those who **abstained** expressed that the task would not be possible for Data Stewards due to resource limitations. Those who were **against** explained that it was an inefficient use of the Data Stewards' time. There was also concern that the monitoring of research was more the duty of researchers and feared there may be overreach if the Data Stewards were responsible for the monitoring.

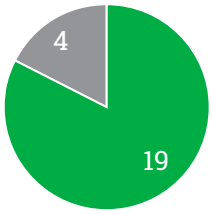
Researchers

15. Researchers have some responsibility to vulnerable populations they study or identify as vulnerable in their research.



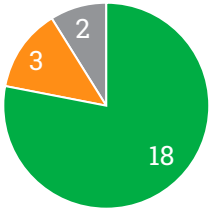
This recommendation was an extension of recommendation 9. Participants were concerned that research results may inadvertently harm the populations they study, particularly vulnerable and/or marginalized populations (e.g., children, First Nations communities). In addition to overview by an independent ethics committee, participants wished to highlight the responsibility of researchers.

16. Anyone seeking access to linked data must sign a standardized contract outlining confidentiality requirements and further dissemination of data.



Participants were concerned that those who were gaining access to linked data were not being adequately informed of their confidentiality requirements and their restrictions on the dissemination of data. They were also concerned about whether there were agreements in place to outline the behaviour of the researcher with possible consequences in cases of breaches with the agreement. Those who **abstained** agreed with the spirit of the recommendation but had an issue with the timing of the signature; they explained that the signature should happen when access is approved.

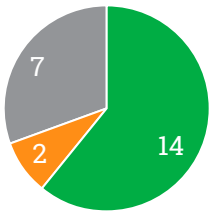
17. Data security certificate program should be established and it should be mandatory for people who are using linked data.



This recommendation was related to the discussion that formed recommendation 12 and 13. Participants were concerned that those who were using linked data may not be adequately trained, thus raising the risk of inappropriate use and security issues. Those who **abstained** were concerned that it was not enough to be certified and that there needed to be a separate auditing process to ensure compliance. Those who were **against** believed the certification was redundant because Data Stewards are already vetting researchers.

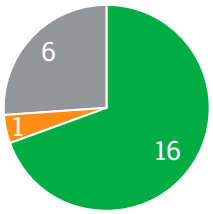
Recommendations regarding public involvement

18. There should be public disclosure (e.g. on a website) of requests for access to data. This should include approvals, denials, and reasons for those decisions.



Participants were concerned about a lack of transparency around when linked data are shared and with whom. They expressed that since there was a lack of transparency, they would not know what types of linked data were being shared, and it would raise concerns that inappropriate uses of data may be occurring without public knowledge. They suggested that transparency in the data access request process may be satisfied by a website displaying the data access requests. Those who **abstained** agreed that there should be disclosure but there is also potential for misinterpretation, and that the disclosure itself could raise concerns and so should be approached cautiously. The group that was **against** said this is not useful because most people would not care. They also said that including the denials would not be necessary, as those requests are usually negotiated, i.e. an initial "no" could become "yes" if the researcher made changes to the request.

19. Transparency and disclosure of research requests is sufficient as a form of public consultation



This recommendation was developed in the context of discussing how the public should be involved in data access requests. Participants reaffirmed the role of transparency and disclosure in sharing knowledge with the public. However, they did not converge on an opinion about how the public should be involved in the data access process. Those who **abstained** were concerned that the term "transparency" was too vague. They were also concerned that only requiring transparency and disclosure would create too low of a bar for government when engaging with the public. The participant who was **against** was also concerned that transparency and disclosure alone would be too low of a bar for the government.

Conclusions

Research from the analysis of linked data can be in the public interest,[15,16] but the use of linked data raises important concerns around the governance, security, and privacy. These concerns need to be addressed while considering the diversity in the perspectives and interests of the public in order to develop policies, procedures, and guidelines that balance the benefits of the research with their risks.[17] Doing so will help ensure that the data sharing process is safe, trustworthy, and acceptable to the public. The importance of working with public can be highlighted in cases where the development of data sharing procedures did not meet the public's expectations and data sharing and access needed to be greatly modified (e.g., care data in the UK[18,19]). Working with the public can take various forms, from surveys, focus groups, to public consultations, and each form will have its benefits and drawbacks.[20]

The public deliberation event, *Using Data About You for Research: Who, When, and How?*, is distinctive from other forms of public engagements in that it was designed and planned to be a forum for members of the public to be informed about and have an opportunity to discuss both benefits and concerns around using linked data. It was a forum where members of the public could engage with each other, share diverse perspectives, and work towards collective conclusions. The event facilitated the considered and informed discussions of a diverse group of British Columbians, and thus the recommendations it produced regarding research using linked data have democratic legitimacy because:

- The recommendations are created by the participants following considered reflections of a diverse group of British Columbians. They are not a snapshot of the opinions of the general public (e.g., as with a survey) and are more in-depth than focus groups.[7]
- Participants were made aware of current policies and controversies regarding the use of linked data for research.
- In cases where there were disagreements on the recommendations, we made sure to record a clear articulation of the disagreement as well as the reasoning behind it, since disagreements and reasons for them can reveal differences in acceptable trade-offs among the public.

The following is a discussion of the deliberation event's main results, outlining how the recommendations support some current practices, suggest changes in guidelines and training, and indicate areas that require further discussion. It is important to emphasize that this is a summary of raw data (the recommendations and stated reasons) with minimal interpretation. A more detailed analysis including analysis of transcripts will contextualize the information within the dynamics of dialogic exchange.[21]

Confirmation of current practices

Participants agreed on the importance of research using linked data and the value it provides to society. Though they were concerned about their privacy and the security of their data, they supported current research activities using linked data. The participants developed ten recommendations that matched current research data access practices in BC, at least for those projects that are managed by Population Data BC (a full comparison of the recommendations and current practices at Population Data BC is available in the Appendix D). The deliberation was set up as a discussion of what should be in place not an evaluation of what is in place, so

it was not a surprise that some recommendations were consistent with current practice. It was noted, however, that some participants struggled with supporting recommendations for practices that they believed already existed. This is in part due to the difficulty in communicating the complexities of the current data sharing regulation environment, but also reflects some feeling that such recommendations would be redundant..

Notably, the deliberation occurred during a period of heavy media coverage of the Cambridge Analytica/Facebook allegations of inappropriate data usage[4]. Prior to the event, the research team discussed the possibility that participants would want to focus exclusively on corporate practices and the resulting privacy issues or would be wary about supporting research using linked data in general. During the deliberation, however, participants all agreed on the importance of research using linked data and the value it provides to society. Generally, they expressed a strong desire to see the data access request process made more efficient in order to facilitate research.

Participants were unanimous in wanting research proposals to be reviewed in order to assess the feasibility of the research and also the ethics of the proposed project. They recognized that reviewing the research was currently done by Data Stewards and approved of this role, though made some recommendations around the codification of the process as further described below.

Participants expressed concern around the potential that research results may have negative impacts as a result of misinterpretation or misuse. They supported the current system of peer review to review research results and publications based on linked data. The participants considered whether an independent party should review for science or to ensure use according to original purpose, but these recommendations, once framed, were not supported by the majority of participants.

For the practices involving researchers, there was strong support for requiring researchers to sign contracts that outline their confidentiality requirements and their responsibility in the dissemination of the data. This conforms with current practices when researchers are granted access to linked data.

Opportunities for change

Participants made a number of recommendations that could be implemented to improve data security, storage, and the efficiency of processing data access requests. The recommendations addressed infrastructure, practice guidelines, public engagement, and training for both Data Stewards and researchers. They generally underlined the need to maintain the security and privacy of data, but also the importance toof increasinge the efficiency of processing data access requests in order to facilitate future research.

Based on concerns related to data security, there was support to establish best practices and guidelines for the storage and access to linked data. There was also support for increased investment to develop linked data sets that could be used for research.

To improve the efficiency of processing data access requests, participants recommended that Data Stewards should all have standard training in order to process the request consistently. They also recommended that the Data Stewards' activities follow standardized policies and procedures. They suggested establishing a certifying process to train the Data Stewards.

Participants did not see the need for policy makers to establish categories to fast-track certain types of research. One reason for this is that they expected that policy makers would fast-track requests to address emergencies anyway (e.g., opioid crisis), so the categories would be redundant. They also expressed that they believed that fast-tracking certain types of research was already something that is being done by Data Stewards.

Participants recommended that researchers undergo a certifying process on how to use linked data securely. They were unanimous in recognizing that researchers have a responsibility to the vulnerable populations they study.

In terms of involving and interacting with the public, participants recommended public disclosure about requests for linked data. There was a strong desire for transparency in the data access request process which may be satisfied by a website displaying the data access requests.

Areas of uncertainty

Participants were divided on some recommendations, sometimes because of wording and other times because of intent. In terms of general oversight of the data access, participants were divided on the need to establish a committee that would monitor all aspects of data sharing, including monitoring the Data Stewards themselves. While the concept received general support, there was concern regarding the breadth of the committee's responsibilities, and that the committee's powers may be misused for political ends.

Participants also diverged on how private enterprises, specifically corporations, should be involved in research. Participants were wary of the intentions of corporations in conducting research (i.e., the profit motive), but many considered that their involvement was inevitable. There was discussion on limiting the corporations' involvement in the research and the benefits they receive, but this was tempered by the fact that it was unlikely that corporations would invest in research without the opportunity to benefit from it. Further, there was a concern that if corporations were limited they would no longer collaborate with researchers.

Concluding thoughts

This public deliberation event produced 19 recommendations for consideration by policy-makers. These recommendations were the result of four days of information sharing and rich discussion in both small and large groups.

This report provides a summary of those recommendations and the reasoning provided explicitly through facilitated discussion following voting. Further analysis will provide greater insight on the process of the deliberations and the context around specific recommendations. Those analyses will offer further considerations for policy and practice in governance of access to linked data.

References

1. Pisani, E. *et al*. Beyond open data: realising the health benefits of sharing data: Table 1. *BMJ Br. Med. J.* **5295**, i5295 (2016).
2. Holm, S. & Ploug, T. Big Data and Health Research—The Governance Challenges in a Mixed Data Economy. *J. Bioeth. Inq.* 1–11 (2017). doi:10.1007/s11673-017-9810-0
3. Lipworth, W., Mason, P. H. & Kerridge, I. Ethics and Epistemology of Big Data. *J. Bioeth. Inq.* 1–4 (2017). doi:10.1007/s11673-017-9815-8
4. http://www.cbc.ca/news/world/facebook-alert-users-data-breach-cambridge-analytica-1.4608621. *CBC Online* (2018).
5. Goodin, R. E. & Dryzek, J. S. Deliberative Impacts: The Macro-Political Uptake of Mini-Publics. *Polit. Soc.* **34**, 219–244 (2006).
6. Burgess, M. M. From 'trust us' to participatory governance: Deliberative publics and science policy. *Public Underst. Sci*. **23**, 48–52 (2014).
7. Solomon, S. & Abelson, J. Why and when should we use public deliberation? *Hastings Cent. Rep*. **42**, 17–20 (2012).
8. O'Doherty, K., Gauvin, F. P., Grogan, C. & Friedman, W. Implementing a public deliberative forum. *Hastings Cent. Rep*. **42**, 20–23 (2012).
9. Longstaff, H. & Burgess, M. M. Recruiting for representation in public deliberation on the ethics of biobanks. *Public Underst. Sci*. **19**, 212–224 (2010).
10. https://www.popdata.bc.ca/bcdatadeliberation. (2018).
11. Burgess, M., O'Doherty, K. & Secko, D. Biobanking in British Columbia: discussions of the future of personalized medicine through deliberative public engagement. *Per. Med.* **5**, 285–296 (2008).
12. O'Doherty, K. C., MacKenzie, M. K., Badulescu, D. & Burgess, M. M. Explosives, genomics, and the environment: Conducting public deliberation on topics of complex science and social controversy. *SAGE Open* **3**, 1–17 (2013).
13. Bentley, C. *et al*. Trade-offs , fairness , and funding for cancer drugs : key findings from a deliberative public engagement event in British. 1–13 (2018).
14. Dry, S. M. *et al*. Community recommendations on biobank governance: Results from a deliberative community engagement in California. *PLoS One* **12**, 1–14 (2017).
15. Brook, E. L., Rosman, D. L. & Holman, C. D. A. J. Public good through data linkage: Measuring research outputs from the Western Australian Data Linkage System. *Aust. N. Z. J. Public Health* **32**, 19–23 (2008).
16. Chamberlayne, R. et al. Creating a population-based linked health database: A new resource for health services research. *Can. J. Public Heal*. **89**, 270–273 (1998).
17. Aitken, M., de St. Jorre, J., Pagliari, C., Jepson, R. & Cunningham-Burley, S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Med. Ethics* **17**, 73 (2016).
18. Sterckx, S., Rakic, V., Cockbain, J. & Borry, P. "You hoped we would sleep walk into accepting the collection of our data": controversies surrounding the UK care.data scheme and their wider relevance for biomedical research. *Med. Heal. Care Philos*. **19**, 177–190 (2016).
19. Vezyridis, P. & Timmons, S. Understanding the care.data conundrum: New information flows for economic growth. *Big Data Soc*. **4**, 205395171668849 (2017).
20. Rowe, G. & Frewer, L. J. A Typology of Public Engagement Mechanisms. Sci. *Technol. Hum. Values* **30**, 251–290 (2005).
21. O'Doherty, K. Synthesising the outputs of deliberation: Extracting meaningful results from a public forum. *J. Public Delib*. **9**, (2013).

Appendix

A. Participant demographic composition

Gender	Age	Income	Ethnic identity	Geographic location
Male: 16	18-24: 4	$20,000-$34,999: 4	African American: 1	Vancouver Coastal: 6
Female: 12	25-34: 7	$35,000-$49,999: 3	Caucasian: 15	Fraser Valley: 7
	35-49: 6	$50,000-$79,999: 9	Chinese: 7	Vancouver Island: 7
	50-64: 7	$80,000+: 12	Egyptian: 1	Northern BC: 4
	65+: 4		Metis: 1	Interior BC: 4
			South Asian: 3	

B. Event Schedule

Agenda for April 7 & 8 and April 21 & 22

DAY ONE: April 7th

8:00-9:00 AM	Breakfast and check-in
9:00-9:15 AM	Welcome address
9:15-10:00 AM	Participant and research team introductions
10:00-10:20 AM	Overview of the event and ground rules
10:20-10:40 AM	Break
10:40-11:00 AM	Speaker 1: Laura MacDougall, BC Center for Disease Control
11:00-11:20 AM	Speaker 2: Ryan Woods, BC Cancer Agency
11:20-11:40 AM	Speaker 3: Michael Vonn, BC Civil Liberties Association
11:40-12:00 PM	Speaker 4: Teri Thorson, Spinal Cord Injury BC
12:00-12:20 PM	Speaker 5: Sherri Pooyak, Aboriginal HIV/AIDS Centre
12:20-1:30 PM	Lunch
1:30-2:30 PM	Speaker panel discussion
2:30-2:40 PM	Introduction to Hopes and Concerns task and break-down of small groups
2:40-3:00 PM	Break (and reconvene in small groups)
3:00-4:00 PM	Small group discussions: Hopes and Concerns
4:00-4:45 PM	Large group discussions: Hopes and Concerns
4:45-5:00 PM	Overview of tasks and goals for Day 2

DAY TWO: April 8th

8:00-9:00 AM	Breakfast and check-in
9:00-9:15 AM	Overview of tasks and goals for the day and introduction of deliberation question #1
9:15-10:15 AM	Small groups: Deliberation question #1
10:15-10:35 AM	Break
10:35-11:55 AM	Large group: Deliberation question #1
11:55-12:00 AM	Introduction to Deliberation question #2
12:00-1:00 PM	Lunch
1:00-2:00 PM	Small groups: Deliberation question #2
2:00-2:20 PM	Break
2:20-3:30 PM	Large group: Deliberation question #2
3:30-4:00 PM	Are there questions we need to add to our agenda for Weekend 2?
4:00-4:30 PM	Overview of tasks and goals for weekend 2, check out

DAY THREE: April 21th

8:00-9:00 AM	Breakfast and check-in
9:00-9:15 AM	Welcome back and overview of weekend
9:15-10:00 AM	Report back on questions from last weekend
10:00-10:20 AM	Break
10:20-11:20 AM	Small groups: Working with scenarios and trade-offs: applying the discussions from the previous weekend
11:20-12:30 PM	Large group: Working with scenarios and trade-offs
12:30-1:30 PM	Lunch
1:30-1:45 PM	Introduction of Deliberation question #3: What processes would make the assessments of risks and benefits from the use of linked data more trustworthy?
1:45-2:45 PM	Small groups: Deliberation question #3 – What processes would make the assessments of risks and benefits from the use of linked data more trustworthy?
2:45-3:05 PM	Break
3:05-4:05 PM	Large group: Deliberation question #3 - What processes would make the assessments of risks and benefits from the use of linked data more trustworthy?
4:05-4:30 PM	Large group: Finalize questions for last deliberation session
4:30-5:00 PM	Overview of tasks for Day 4

DAY FOUR: April 22th

8:00-9:00 AM	Breakfast and check-in
9:00-9:15 AM	Welcome back and overview of day
9:15-10:15 AM	Large group: Participants each share the issue and/or concern they most want addressed regarding the sharing of linked data
10:15-10:35 AM	Break
10:35-11:45 PM	Large group: Development of policy recommendations and voting
11:45-12:45 PM	Lunch
12:45-1:45 PM	Large group: Review and revise recommendations
1:45-2:45 PM	Expert and policy panel discussion
2:45-3:00 PM	Break
3:00-4:00 PM	Large group: Considerations from policy panel discussions
4:00-4:30 PM	Wrap up, check-out, and thank you!

C. Data access request scenarios used for discussing trade-offs

SCENARIO 1

What is the research question?
What are the health and education outcomes for refugee children who landed in BC between 2005 and 2009?

Who is asking for data access?
A student working on a PhD thesis

What data are being requested?
Immigration, hospital separations, physician payment data, prescription drugs, Early Development Instrument, education data on testing and school leaving, plus demographic including family size, number of sibling, and neighbourhood average income

How is individual identity protected?
Asking for birth year only (not full birth date), region of province (16 regions) not city or neighbourhood, asking for region of immigration not specific country

Who is interested in the research? Whose priority is this?
The student and her supervisor

Who is going to be involved in the research?
The student, her supervisor, and two other committee members who are academics at BC universities

Who is funding the research?
There is no specific funding for this research

Where will the data be stored / analyzed?
The data will be stored and analyzed in a known and government-approved secure research environment

Is there an ethics review in place?
Yes, University ethics is in place

Is there proof of scientific review?
The supervising committee provides assurance of scientific review

SCENARIO 2

What is the research question?
What are the social and health predictors of having either a fatal or non-fatal opioid overdose?

Who is asking for data access?
The BC Centre for Disease Control

What data are being requested?
Early Development Instrument, education testing and school leaving, WorkSafeBC (workplace injury), physician payment data, hospital separations, prescription drugs, neighbourhood of residence (and changes in that over time), income (from PharmaNet registration) and changes in that over time

How is individual identity protected?
Asking for birth year only (not day and month)

Who is interested in the research? Whose priority is this?
The Provincial Health Officer asked for this, related to an opioid overdose being declared a public health emergency

Who is going to be involved in the research?
BC Centre for Disease Control scientists, current and former drug users, the Provincial Health Officer, PhD students

Who is funding the research?
The Ministry of Health

Where will the data be stored / analyzed?
In a secure environment at the BC Centre for Disease Control

Is there an ethics review in place?
Yes

Is there proof of scientific review?
The study was reviewed and approved by the Provincial Health Officer

SCENARIO 3

What is the research question?
What are the employment experiences of Indigenous people who move off reserve?

Who is asking for data access?
A University-based researcher

What data are being requested?
Indigenous status information, WorkSafeBC (workplace injury), social assistance, payer of premiums for health care (can indicate employer), Vital Statistics births and marriages data (to identify spouse and children), neighbourhood average income (and changes in that over time)

How is individual identity protected?
Full birth date requested, and community of residence

Who is interested in the research? Whose priority is this?
Industry Association – in relation to employee training programs

Who is going to be involved in the research?
Academic researcher and representatives of the Industry Association

Who is funding the research?
Industry Association

Where will the data be stored / analyzed?
At the university academic's office

Is there an ethics review in place?
Not yet

Is there proof of scientific review?
This was reviewed by the Industry Association

D. Population Data BC current practices and recommendations

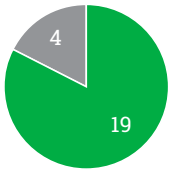
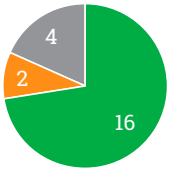
Below, we outline the recommendations from the deliberation event compared to current practices at Population Data BC. We also include additional explanatory comments and the pie chart that shows the votes and the level of support for each recommendation. Please refer to the main text for details on the voting and the participant reasoning behind the recommendations.

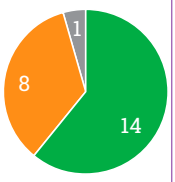
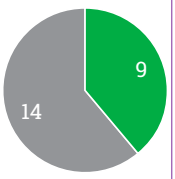
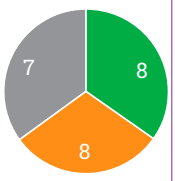
Population Data BC is an organization that links data and provides supports for researchers to request access to those data. It is important to note that PopData supports academic research requests, but not all requests are for access to linked data. As the organization organizing the public deliberation it is important for PopData to assess the recommendations against current practice. This should not be interpreted as how all data requests are handled in BC.

As in the main document, the participant recommendations are grouped into four categories (they are not in order as they occurred in the event): 1. Recommendations regarding the governance of linked data; 2. Recommendations regarding the security and review process for linked data and the results from their analysis; 3. Recommendations regarding the responsibilities of researchers and Data Stewards; 4. Recommendations regarding public involvement.

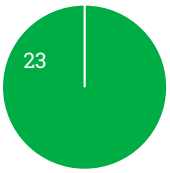
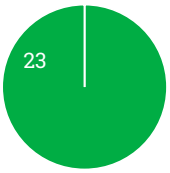
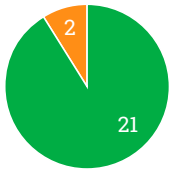
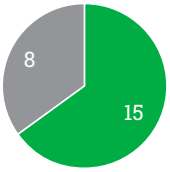
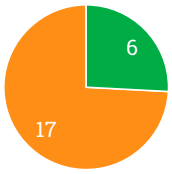
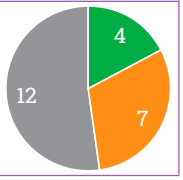
In each section, the recommendations are ordered by the degree to which the participants converged in their positions. In some cases, there was not strong convergence. The reasons given by some participants when voting against or abstaining were sometimes but not always about opposition to the substance of the recommendation. The reasoning for votes cast sometimes suggested agreement in principle but disagreement with the wording. In other cases, reasoning revolved around whether a recommendation was necessary if it reflected current practice.

Recommendations regarding the governance of linked data

Recommendation	PopData current practice	Comments	Results
1. Develop a plan to make the data linkage approval process more efficient, without compromising the evaluation process.	PopData has been working with Data Stewards to promote the idea of "proportionate governance" and through that to increase transparency around approval processes. The Ministry of Health has been leading a process for streamlining access to Ministry and Health Authority data, and PopData has been participating in that.	Continue with current plans, and ensure the results are documented and made transparent to the public.	
2. It is important to invest in a collection of linked datasets to promote efficient research while enhancing privacy protection.	PopData was created to address the difficulty and inefficiency of linking data on a project-by-project basis, but without creating a "giant database" of linked information.	Abstentions and disagreements suggest that it is important to maintain the current practice of keeping data separate and unavailable unless or until there is an approved use.	

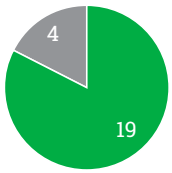
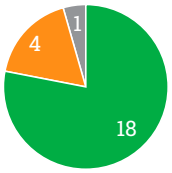
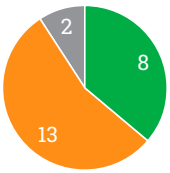
Recommendation	PopData current practice	Comments	Results
3. Policy makers should establish categories that identify requests that require different paths or speed for review, e.g., fast track for urgent research priorities.	PopData has been promoting a method of ranking data access requests by their level of risk (i.e., proportionate governance) and reviewing accordingly.	Some Data Stewards have been implementing this method tailored to their needs. PopData will bring to Data Stewards for further discussion.	
4. If a commercial entity funds research with linked data they should not be involved in the production and review of that research.	Commercial entities are not currently eligible to request access to data through PopData.	Participants did not converge in voting on this recommendation. This is an area worthy of further discussion and further analysis of the deliberation's results.	
5. There should be a committee or governing body with authority to: 1. Provide oversight and investigation for breaches and/or harms 2. Apply penalties or other consequences 3. Develop policies to mitigate the potential for future breaches and/or harms 4. Intervene when Data Stewards disagree 5. Develop and operate an appeals process 6. Provide certification for Data Stewards	PopData currently supports a policy-setting Data Stewards Working Group. This group, however, does not have a formal appeals process or certification for Data Stewards.	Participants did not converge in voting on this recommendation. PopData to take this recommendation to the Data Stewards Working Group and others for discussion.	

Recommendations regarding the security and review process for linked data and the results from their research

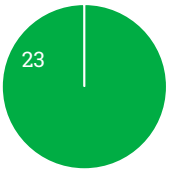
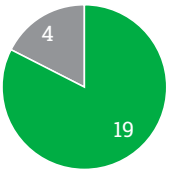
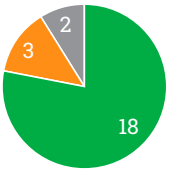
Recommendation	PopData current practice	Comments	Results
6. Scientific review of the research proposal should be performed by an independent party.	Scientific merit is determined by research funding agencies, and proof of scientific merit is a requirement for application for data through PopData. Graduate student projects are considered to have peer review if they have a letter of support from the thesis supervisor. PopData has an established proxy peer review process for applications that have not received external review and approval.		
7. There should be best practices and guidelines for secure storage and access to linked data.	PopData has established rigorous practices and guidelines for secure storage of data. PopData operates a secure virtual environment for data analysis. PopData also offers education and training on how to use linked data.	This is an area to continue to develop best practices and make these more transparent.	
8. Results and publications of linked data research must be reviewed to ensure that they are justified by the analysis of the data.	PopData requires researchers to provide for review any publication that is intended to be made public. The review is done both by PopData and Data Stewards.		
9. The proposed research and data access should be reviewed by an independent ethics committee to ensure benefits outweigh potential harms (e.g. potential for re-identification, stigma)	Proof of ethics review is a requirement for application for data access through PopData.		
10. Research results should be reviewed by a qualified independent party to reaffirm the original purpose of the research.	PopData requires researchers to justify fit of research results with the original request when those materials are submitted for Data Steward review.	Public deliberation suggests this approach may be beyond what is necessary. This warrants further discussion.	
11. An independent party should assess requests for data to be sure that the data are necessary to conduct the research.	PopData currently helps researchers request the data that matches the needs of the research proposal. Data Stewards confirm that fit.	Participants did not converge in voting on this recommendation.	

Recommendations regarding the responsibilities of researchers and Data Stewards

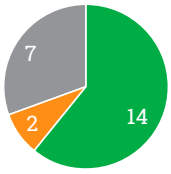
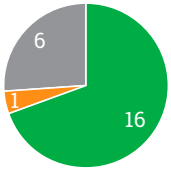
Data Stewards

Recommendation	PopData current practice	Comments	Results
12. Data Stewards should have standard training or certification to ensure appropriate expertise for their role.	There is currently no training for Data Stewards, though PopData has a Data Steward Working Group which provides a forum for discussion and policy-setting across Data Stewards.	This requires further discussion.	
13. Data Stewards should have standard policies and procedures to guide their work and there should be a certifying body to maintain those.		There is currently no such body.	
14. Research using linked data must be monitored by Data Stewards to ensure data are used in accordance with the original request.	Data Stewards require that all research products are reviewed prior to being made public. PopData manages this process.	Participants did not converge in voting on this recommendation. Further discussion is required.	

Researchers

Recommendation	PopData current practice	Comments	Results
15. Researchers have some responsibility to vulnerable populations they study or identify as vulnerable in their research.	PopData processes require ethics review and approval for data access requests. This issue should be covered by that ethics review.	This is currently covered under the Tri-Council Policy Statement on best practices for research involving human subjects, which governs all University-based research.	
16. Anyone seeking access to linked data must sign a standardized contract outlining confidentiality requirements and further dissemination of data.	PopData requires that the researchers sign a Research Agreement before they obtain approved data. Researchers also must pass specific privacy training before getting access to data and using the PopData Secure Research Environment.		
17. Data security certificate program should be established and it should be mandatory for people who are using linked data.	PopData currently has privacy training for researchers, but not as extensive as a certificate program or researcher accreditation program. PopData provides researchers access to data mainly through a secure virtual environment, which means that researchers do not require expertise in data security systems.		

Recommendations regarding public involvement

Recommendation	PopData current practice	Comments	Results
18. There should be public disclosure (e.g. on a website) of requests for access to data. This should include approvals, denials, and reasons for those decisions.	The PopData website discloses information for approved research projects and writes more in-depth case studies for a subset of those projects.	PopData does not share projects that have not been approved.	
19. Transparency and disclosure of research requests is sufficient as a form of public consultation.	PopData is working to develop a sustained means of public engagement and consultation.	This would benefit from further discussion.	

Report authors:

Jack Teng
Colene Bentley
Michael Burgess
Kieran O'Doherty
Kimberlyn McGrail

Acknowledgements:

We would like to acknowledge the funding support from the Canadian Institutes for Health Research (CIHR). We would also like to acknowledge the expert feedback from the project collaborators: Holly Longstaff, Matt Reed, Jeannette van den Bulk, Suzanne Vercauteren, and Sue Wheatley.

We also thank the small group facilitators: George Athans, Samantha Pollard, Darielle Talarico, and Jessica White.

June 2018
Vancouver, British Columbia, Canada