murdoch children's research institute

Open conversations about Open Data: Planning for data sharing in LifeCourse Cohorts

Last updated: 3.10.2022

Version: 1.0

Acknowledgements

LifeCourse acknowledges all collaborators, cohort representatives and participants. See https://lifecourse.melbournechildrens.com/contact/ for further details.

Suggested citation

O'Connor, M., on behalf of the LifeCourse Initiative. 2022. Open conversations about Open Data: Planning for data sharing in LifeCourse cohorts. Murdoch Children's Research Institute. Melbourne, Australia. https://doi.org/10.25374/MCRI.21266586.

Contents

Background

Each of the LifeCourse cohorts hold a wealth of valuable data that can be used to advance child and adolescent health. Appropriate reuse of these data is not only cost-efficient, given the large investments already made in data collection, but also ethical, with no additional burden on participants.

Increasingly, it is expected by journals and funders that data will be appropriately shared, and this is valued and counted through metrics like dataset citations. The National Statement on Ethical Conduct in Human Research states that:

> *"In the absence of justifiable ethical reasons (such as respect for cultural ownership or unmanageable risks to the privacy of research participants) and to promote access to the benefits of research, researchers should collect and store data or information generated by research projects in such a way that they can be used in future research projects."*

This shift towards 'Open Data' is part of a broader push to implement open science practices that promote efficient, transparent, and reproducible research.

Of course, we want to make sure that data sharing is secure and appropriate because we care deeply about the wellbeing and privacy of our participants. It is also important to manage potential organisational risks (find out more about data breaches here).

Data sharing should align with relevant MCRI policies and procedures. It must also align with what participants agreed to at the time they provided consent, or with ethically and legally acceptable uses beyond this, approved by the relevant HREC. LifeCourse cohort studies are often undertaken in collaboration with other organisations, who may have additional requirements.

A range of supports are available to assist your cohort team in developing an appropriate data sharing plan:

- For questions about ethical considerations relating to data sharing, RCH Ethics is here to help
- LifeCourse provides a common online data access application form, acting as an intermediary while cohort custodians decide on whether to enter a new collaboration
- For assistance with organisational agreements and other legal matters, contact MCRI Legal
- Data sharing should be planned in the context of a broader data management approach; find relevant templates and resources here
- Find links to these and other resources on the LifeCourse website and in our LifeCourse Data Access Hub for cohort custodians

This document is intended as a resource to facilitate open and pragmatic discussion amongst LifeCourse cohort teams during the development of their data sharing plans, with tips on where to find further information. There may be other questions relevant for your cohort to consider beyond those listed, and it is worthwhile investing the time in a full and open discussion about data sharing at the outset of your cohort's development.

Given their long running nature, LifeCourse cohorts also benefit from periodically reviewing and updating their data sharing processes. Similarly, standards and organisational policies continually evolve, and this will be reflected in ongoing updates to this resource. Share any feedback with us at lifecourse@mcri.edu.au.

Discussion prompts

About your cohort's data sharing intentions

1. Do you plan to share research data beyond the study team where appropriate and in line with ethics and governance procedures?

If yes, why? Document your motivation for this decision, such as the value of the data collected and capacity for these data to be used in wide ranging ways that benefit child and adolescent health. Describe the conditions (e.g., participant consent for future research use) that further justify sharing from an ethical, legal, and privacy perspective.

If no, why? Documenting your rationale for this decision is recommended as this will likely need to be justified to funders and journals in future. It is increasingly expected that appropriate data sharing processes be in place, in line with the National Statement on Ethical Conduct in Human Research.

About your cohort's data sharing process

2. How will you ask researchers to apply for data access?

LifeCourse provides a common online data access application form for LifeCourse cohorts: https://lifecourse.melbournechildrens.com/data-access/. Consider directing interested data users to complete their application via LifeCourse, to reduce your administrative load while remaining in full control of all data access decisions relating to your cohort (read more about governance here and mechanics here).

3. What criteria will you use to evaluate data access applications?

Examples of common criteria used by cohorts may include:

- Suitability of the proposed project given the available data;
- Alignment of the project to the purpose and strategic plans of the cohort;
- Redundancy with other planned or in progress work; and
- Scientific quality of the proposed work.

4. How will you ensure that the data user has ethical approval for their analysis?

Even when not directly collecting data, ethics approval is needed to work with and publish from research data (see here). External data users who are not part of your cohort's study team and are using the data in a way that extends beyond the study's original aims may need to submit a new ethics project covering their secondary data analysis project. This is the responsibility of the data user to progress.

Cohorts require a process to verify that this ethical approval is in place. For example, a data user may be required to provide their letter of ethics approval and a copy of the protocol prior to data release. An example of a low/negligible risk ethics protocol covering analysis of existing data can be found here from the LifeCourse-COVID-Wellbeing research program.

5. What data will be shared?

Different types of data (e.g., highly sensitive data like personal identifiers, sensitive data such as health information, and protected data like anonymised information) have different levels of risk for re-

identification of participants, and warrant different approaches to sharing. See the MCRI Research Data Classification Policy for details, accessible by contacting dataoffice@mcri.edu.au. Efficiencies may be gained from defining lower-risk deidentified or anonymised dataset/s that can be shared to meet the needs of most analyses, while highly sensitive variables are only provided on an as-needed-and-approved basis.

6. For which participants will data be shared?

The use of data should be in line with the information provided to the participant about how their data would be used at the time of their consent (e.g., as detailed in the PICF/PGIS), and to which they have agreed. This may require, for example, removing those participants who did not consent to future research use from dataset/s prior to sharing.

Sometimes there is a strong rationale for sharing and using the data in a way that does not align with what participants had initially consented to. An example of this is where participants were not asked about consent for future research use, and it is no longer practicable to gain this consent. In this case, researchers may provide an argument as part of their ethics application for why extending the use of data beyond the original terms of consent is ethically acceptable. Alternatively, further consent from participants can be gained by updating participants about changes to how their data will be utilised. For more information contact RCH Ethics, and for assistance with legal or privacy issues contact MCRI Legal.

7. What steps will be taken to reduce the risk of participants being reidentified?

Cohorts may need to undertake additional checks to reduce the risk of re-identification of participants within the dataset/s to be shared. This should include checking that no variables contain unique values that could directly, or in combination with other variables in this or other shared datasets, pose a re-identification risk. A data de-identification and anonymisation Standard Operating Procedure (SOP) is currently in development to assist with this process; contact the MCRI Data Office (dataoffice@mcri.edu.au) for further information.

8. Who will you make the data available to?

Define who will be provided with access to the data. For example, this may specify investigators who have been named on the secondary data analysis ethics protocol and who will be undertaking data analysis. You may choose to implement vetting procedure to ensure data are entrusted with legitimate researchers from recognised academic institutions in Australia and/or overseas.

9. In what format will the data be shared?

Outline what program format/s the data file will be transferred as, such as STATA, R, csv, or other file formats. Another consideration is how the data file/s will be structured, such as separate data files relating to each wave of data collection.

10. How will data be made available?

Find out more about approved MCRI IT systems here, and recommended options for transfer of files containing research data here. For example, data can be securely transferred via REDCap Send-It (more information here).

11. What study documentation will be provided to data users?

Rich descriptions of cohort data are critical to making it meaningful and reusable by others. Relevant data documentation to provide might include: the study protocol, Data dictionary, Data user guide, link to your cohort's entry in the LifeCourse metadata catalogue, and other useful documentation for understanding the study context and design.

12. What expectations will approved data users be required to agree to in your Data User Agreement or similar?

A Data User Agreement is an opportunity to explicitly detail expectations before the project begins, such as requirements around data security, privacy, return of data, authorship and acknowledgements. Depending on the nature of the relationship, an alternative agreement such as a Research Collaboration Agreement or a Data Sharing Agreement may be recommended to cover the data use and any other requirements (see #13). For assistance with agreements, contact MCRI Legal.

13. How will your cohort arrange organisational agreements when required and ensure that a valid agreement is in place?

MCRI Legal helps study teams to form appropriate legal agreements with external organisations when data users are external to MCRI. For more information about different types of organisational agreements see here. What process will your cohort use to notify the legal team when an agreement is required, and to keep track of these agreements? For assistance with agreements, contact MCRI Legal.

Documenting and communicating your data sharing plan

14. Are details of your cohort's approach to data sharing included in your study's approved ethics protocol?

Details of your data sharing process such as those discussed in the previous section should be reviewed and approved by the relevant HREC. If updating your data sharing process, consider whether an amendment to your ethics protocol is required. It is also recommended to include key expectations in the Data User Agreement or similar, as this ensures that data users are aware of the requirements and provides a legal basis for enforcing them (see #12).

15. What is your cohort's Data Sharing Statement to be included in publications?

These short statements are now required by many journals, to inform interested data users about whether there is potential for data sharing at the current time and if so how to proceed. For example, "Ethics approvals do not permit these potentially re-identifiable data to be made available in a public repository, but access can be requested at https://lifecourse.melbournechildrens.com/data-access/."

16. Are key features of your cohort's data sharing plan transparent for potential data users?

Ideally, details of a study's data sharing process should be publicly available and accessible for potential data users to review prior to deciding whether to lodge a request. Consider opportunities and avenues available to your study to promote transparency. For example, a summary of your process could be displayed on your study website and included in your cohort's protocol paper.

Melbourne Children's LifeCourse Initiative

E: lifecourse@mcri.edu.au
W: https://lifecourse.melbournechildrens.com/

Murdoch Children's Research Institute

The Royal Children's Hospital
50 Flemington Road
Parkville, Victoria, 3052 Australia

murdoch children's research institute

LifeCourse