CanDLe User Protocol

Lead Researcher
Data Controller
Approved Data User

Version 1.3 | 17 June 2021

NSW GOVERNMENT

Cancer Institute NSW

Contents

Definitions

Term	Definition
Approved Data User (ADU)	Lead Researcher, Data Controller, or a member of the CanDLe Lead Researcher's group who is authorised by the CanDLe Co-ordinating Principal Investigator to have access to the CanDLe data under the responsibility of the Lead Researcher.
Community of Practice (CoP)	A group comprised of Lead Researchers and relevant team members from NSW Institutions and the Independent Review Committee to collaborate, share knowledge and conduct peer review of research sub-studies.
Co-ordinating Principal Investigator (CPI)	The CPI takes responsibility for the overarching CanDLe project in accordance with Ethics approval.
Data Controller	A named individual (may be the Lead Researcher) responsible for overseeing (reviewing and approving) all incoming and outgoing data flow and files on the secure data environment. The Data Controller is also responsible for ensuring that any information entering the data environment complies with Ethics and Data Custodian approvals. Two Data Controllers may be nominated, in addition to the lead researcher.
E-Research Institutional Cloud Architecture (ERICA)	A secure cloud computing environment that provides an appropriate level of security for research that uses sensitive microdata.
Independent Review Committee (IRC)	A Committee comprised of appropriately experienced professionals to review and recommend Lead Researchers to the CPI and PHSREC. The IRC will also review and advise on any issues escalated from the CoP.
Lead Researcher	The Lead Researcher/Principal Investigator is responsible for all sub-studies conducted by their research group.
Population Health & Services Research Ethics Committee (PHSREC)	The NSW PHSREC undertakes scientific and ethical review of population health and/or public health research which utilises or links to routinely collected health (and other) state-wide data.
Secure Unified Research Environments (SURE)	A remote-access computing environment that allows researchers to access and analyse linked health-related data files. This is located at the Sax Institute.

About CanDLe

Enduring Cancer Data Linkage (CanDLe) for health services research in New South Wales is a Cancer Institute NSW initiative that will provide linked unit record cancer data to a network of approved researchers, aiming to advance the pace, quantity and quality of population-wide cancer research that is aligned to the goals of the NSW Cancer Plan.

The initiative has been designed to safely provide faster access to linked cancer data and follows the Five Safes framework. Through its emphasis on researcher collaboration, CanDLe serves as an infrastructure to facilitate high quality population health research in NSW.

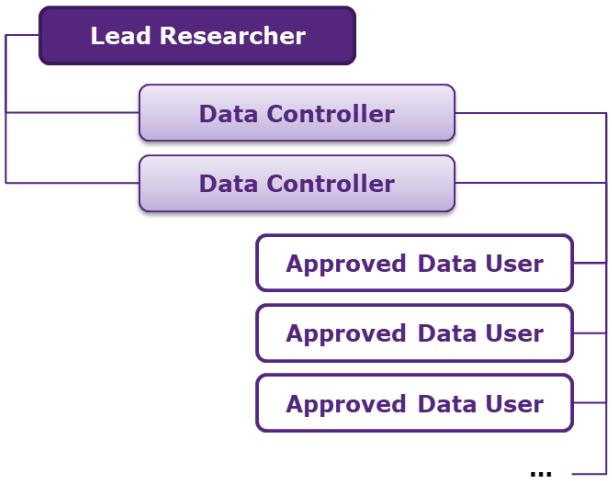
Ethics

The Population & Health Services Research Ethics Committee (PHSREC) has reviewed and approved the CanDLE project (2019/ETH12584). This overarching approval enables the Cancer Institute NSW to create 2 data assets (CanDLe 1, CanDLe 2 Women Screen) including the approved variables. All sub-study protocols submitted by Lead Researchers and recommended by the Community of Practice must be reviewed and approved by the PHSREC prior to commencement of the project.

Users

All researchers using the CanDLe datasets are responsible for ensuring they adhere to the Five Safe's Framework[1,2,3]. There are three tiers of users of the CanDLe data who have specific roles and responsibilities. They are 1) Lead Researcher, 2) Data Controller, and 3) Approved Data User (Figure 1). The Lead Researcher is responsible for ensuring that the roles of the Data Controller are fulfilled and for their research team's appropriate management and use of the CanDLe data. The Data Controller is responsible for overseeing the appropriate access and use of the CanDLe data by the Approved Data Users within their secure data storage workspace. The Approved Data Users are responsible for safely and ethically using the CanDLe data. The roles and responsibilities are outlined in more detail below.

Figure 1. Hierarchy of CanDLe Data Users



[1] Ritchie, F. The Five Safes: A framework for planning, designing and evaluating data access solutions. In: Data for Policy 2017, London, England, 6-7 September 2017. Available from: http://eprints.uwe.ac.uk/36263.

[2] Department of the Prime Minister and Cabinet. New Australian Government Data Sharing and Release Legislation: Issues paper for consultation, 4 July 2018. Canberra: Australian government, 2018. Available from: https://www.pmc.gov.au/resource-centre/public-data/issues-paper-data-sharing-release-legislation.

[3] Young A & Flack F. Recent trends in the use of linked data in Australia. Australian Health Review 2018; 42: 584-90. Available from: https://www.publish.csiro.au/AH/pdf/AH18014.

Lead Researcher

The **Lead Researcher** is responsible for the safe and ethical use of the CanDLe datasets within the secure data storage workspace by their research team. They are responsible for ensuring that the roles and responsibilities of the Data Controller(s) and Approved Data User(s) (ADUs) are followed. They are also responsible for the overall management of all approved sub-studies of their research group, as well as for the ongoing collaboration between the researchers who participate in CanDLe through the Community of Practice (CoP) to increase our understanding of cancer in NSW.

They are responsible for:

Safe People

1) Ensuring that the role of a Data Controller is fulfilled either by themselves or by (up to two) approved nominated individual(s), and to contact the CanDLe secretariat at CINSW-CandleProgram@health.nsw.gov.au to propose any changes to the controller;
2) Ensuring that each member of their research team agrees to and signs the Data Access Declaration prior to accessing the data;
3) Ensuring that each member of their research team completes the required training;
4) Ensuring that data is accessible only by approved data users;
5) Informing the CanDLe secretariat of any change to the ADUs on their research team;
6) Agreeing to the Terms of Reference for the Community of Practice;

Safe Setting

7) Ensuring that their research teams have appropriate institutional facilities (physical space, equipment, software, technology and computing capabilities), data governance policies and procedures and resources to carry out the proposed research;
8) Setting up their team's access to an approved secure data storage workspace and funding any associated costs;

Safe Project

9) Participating on the Community of Practice to collaborate, share knowledge and conduct peer-review of research sub-studies to ensure they are of high quality, aligned to the goals of the NSW Cancer Plan and that there is reduced duplication of research topics;
10) Ensuring all Sub-studies be conducted according to the **NHMRC Australian Code for the Responsible Conduct of Research**;
11) Submitting Sub-Study Protocols to the Community of Practice for peer-review;
    a) The protocol should include a sufficient literature review outlining the research question, study design, statistical analysis plan, and outcome measures to be used.
12) Submitting an **annual progress report** for **each** Sub-Study they are responsible for to the CanDLe secretariat by **31 October** each year. They must also submit a final report once the project has been completed. All outcomes and planned publications/

presentations must be included in this report. Please send these to CINSW-CandleProgram@health.nsw.gov.au;

Safe Data

13) Ensuring that data or any information discovered from the data, may not be used for commercial purposes and may not be redistributed to third parties without written permission from the Co-ordinating Principal Investigator;

14) Reporting any data breaches or protocol violations to the Chair of the IRC within 14 days of becoming aware of the occurrence. Access to data may be revoked for failure to report and based on the severity of the breach or violation;

15) Ensuring Aboriginal Health and Medical Research Council ethics approval has been given prior to the analysis of any data where indigenous status will be reported in the results;

16) Ensuring that all copies of the information derived from this data will be destroyed in accordance with ethical approval and if requested by the Co-ordinating Principal Investigator;

Safe Output

17) Ensuring that if requested, codes used to undertake analyses will be provided to the Community of Practice to ensure research integrity and reproducibility;

18) Ensuring that any draft publication, report or presentation using CanDLe datasets is peer-reviewed by the Community of Practice and the Co-ordinating Principal Investigator (or delegate) prior to submission for publication;

   a) CanDLe Co-ordinating Principal Investigator reserves the right to require that any publication which could potentially breach the conditions of the ethics approval or data custodian conditions of release be withheld from submission for publication.

   b) If there are any significant changes to the paper after submission, re-review from the Co-ordinating Principal Investigator must be sought prior to submitting the changes.

19) Ensuring that research findings are only presented in aggregate form with sufficiently large cell sizes to ensure that no individual can be identified (suppressing cell sizes <5) in peer review publications, conference presentations and the public domain;

Other Responsibilities

20) Acknowledging the NSW Ministry of Health, Cancer Institute NSW and the Centre for Health Record Linkage for any reports or publication using data from the CanDLe Initiative with the following text:

   a) "This research was completed using data from the Cancer Data Linkage (CanDLe), a Cancer Institute NSW initiative version [INSERT VERSION]. The CanDLe Initiative is supported by the NSW Ministry of Health and the Cancer Institute NSW. Record linkage was provided by the Centre for Health Record Linkage."

21) Providing a copy of any publication or report resulting from the use of CanDLe data to the CanDLe secretariat at CINSW-CandleProgram@health.nsw.gov.au **on acceptance** of publication;

22) Providing a **copy of any publication** or report to the NSW Ministry of Health and Cancer Institute NSW at least **two weeks prior to public release**, emailed to moh-cee@health.nsw.gov.au and copy in CINSW-CandleProgram@health.nsw.gov.au;

23) If using the Cause of Death Unit Record File data, acknowledging the Australian Coordinating Registry by using the following:

   a) "The Cause of Death Unit Record File (COD URF) is provided by the Australian Coordinating Registry for the COD URF on behalf of the NSW Registry of Births, Deaths and Marriages, NSW Coroner and the National Coronial Information System."

   b) "Source: Cause of Death Unit Record File held by the NSW Ministry of Health Secure Analytics for Population Health Research and Intelligence."

24) If using the Cause of Death Unit Record File data, sending a copy of any publication or report to the Australian Coordinating Registry for the COD URF **at least two weeks prior to public release**, emailed to BDM.CODURF@justice.qld.gov.au and copy in CINSW-CandleProgram@health.nsw.gov.au.

Data Controller

The Data Controller, who may also be the Lead Researcher, is responsible for overseeing the appropriate access and use of the CanDLe data by the Approved Data Users within their secure data storage workspace.

They are responsible for:

Safe People

25) Having experience or training handling potentially sensitive personal information or re-identifiable data;

26) Contacting the CanDLe Secretariat at CINSW-CandleProgram@health.nsw.gov.au to propose any changes to the data controller;

Safe Setting

27) Overseeing the research team's workspace within the secure data storage workspace;

Safe Project

28) Ensuring that the CanDLe datasets are only used for the research questions and analyses in the approved Sub-Study Protocols;

Safe Data

29) Ensuring that any information entering the workspace complies with Ethics and Data Custodian approvals;

30) Ensuring that data will not be matched with information on individuals from another source other than the datasets specified in the approved ethics application;

Safe Output

31) Reviewing and approving all incoming and outgoing data flow and files on the secure data storage workspace;

32) Never removing data at a unit record level;

33) Ensuring that no more than 10 pages to one Microsoft Word document or 10 tabs to a Microsoft Excel file;

34) Ensuring that documents reporting percentages include frequencies (or 'N's);

35) Ensuring that cell counts should be five (5) or more;
   a) If cell count is less than five (5), you can aggregate variables, e.g., age groups 65-74, 75-84, 85+ are replaced with 65+. If aggregation is not possible, then cells may be suppressed.
   b) If your table includes rows and/or column totals, make sure that at least one other cell in a row and column is also suppressed, or the row and column total is suppressed.

36) Ensuring that you check that:
   a) All columns and rows in Microsoft Excel have been 'unhidden';
   b) Graphs are accompanied by frequency tables;
   c) All columns/rows/variables are clearly labelled;
   d) Log files do not contain any unit record data;
   e) Program files in SAS, Stata, etc. do not contain comments that report small frequencies; and
   f) Zip files do not contain more than 10 files, and no files are nested within files.

Approved Data User

An Approved Data User (ADU) includes the Lead Researcher, Data Controller, or a member of the CanDLe Lead Researcher's group who is authorised by the CanDLe Co-ordinating Principal Investigator to have access to the CanDLe data. A PhD student or early career researcher may be an ADU on sub-studies, if appropriate supervision is provided, e.g., a supervisor is on the Lead Researcher's team and the Lead Researcher is satisfied that the ADU will meet the requirements of the 'Five Safes Model'. All ADUs are also bound by local Institutional data usage policies.

An ADU is responsible for:

Safe People

37) Signing and agreeing to the Data Access Declaration prior to accessing the secure data storage workspace;

38) Undergoing all required training prior to accessing the data;

Safe Setting

39) Ensuring that the CanDLe datasets are only stored on the research team's approved virtual computing environments throughout the duration of this project. The current approved environments are:

   a) Secure Unified Research Environments (SURE), which can be accessed using an encrypted connection from researcher's local computer. The dedicated virtual computing environments runs on hardware centrally managed by the Sax Institute. SURE requires that all researchers 1) undertake training on issues of privacy, ethics, information security and statistical disclosure control prior to gaining access to SURE and 2) sign a deed outlining the terms and conditions of using SURE.

   b) E-Research Institutional Cloud Architecture (ERICA), which can be accessed through remote-access Windows virtual workstations, with anti-virus and malware detection systems. All users are required to undertake training in ethics, privacy, information security, statistical disclosure control and use of ERICA.

Safe Project

40) Ensuring that all sub-studies will be carried out in accordance with the 'Five Safes Model' as outlined in the Summary Protocol;

41) Ensuring that the data will only be used for ethically approved sub-studies;

Safe Data

42) Ensuring that data are not held on any portable devices including external drives, laptops or desktop computers, or transferred to an alternate server;

43) Ensuring that data will not be matched with information on individuals from another source other than the datasets specified in the approved ethics application;

Safe Output

44) Ensuring that no information will be released with which it may be possible to identify an individual person (i.e. unit record data);

45) Ensuring that research findings are only presented in aggregate form with sufficiently large cell sizes to ensure that no individual can be identified (suppressing cell sizes <5) in peer review publications, conference presentations and the public domain.